RL-TR-95-84
In-House Report
August 1995

EXTERNAL/INTERNAL DATA FUSION TESTBED: HISTORY, COMPONENTS AND EXPERIMENTAL ANALYSIS

John G. Parker

DTIC
S ELECTE D
OCT 1 9 1995
F

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

19951017 086

DTIC QUALITY INSPECTED 5

Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-95-84 has been reviewed and is approved for publication.

APPROVED:

JOSEPH CAMERA, Chief
Signals Intelligence Division
Intelligence and Reconnaissance Directorate

FOR THE COMMANDER:

ANTHONY S. SZALKOWSKI
Acting Technical Director
Intelligence and Reconnaissance Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (IRAA) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1995	In-House

4. TITLE AND SUBTITLE
EXTERNAL/INTERNAL DATA FUSION TESTBED: HISTORY, COMPONENTS AND EXPERIMENTAL ANALYSIS

5. FUNDING NUMBERS
PR-3188
TA-07
WU-50

6. AUTHOR(S)
John G. Parker

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Rome Laboratory (IRAA)
32 Hangar Rd
Griffiss AFB, NY 13441-4114

8. PERFORMING ORGANIZATION REPORT NUMBER
RL-TR-95-84

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Laboratory (IRAA)
32 Hangar Rd
Griffiss AFB, NY 13441-4114

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: John G. Parker/IRAA (315)330-4025
A Master's Thesis presented to the SUNY Institute of Technology at Utica/Rome

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
Rome Laboratory (RL) has been performing Speech Processing Research and Development since the early sixties. Areas of research include Speech Recognition, Speaker Identification, Language Identification, Keyword Recognition, Platform Identification, and External/Internal (E/I) Data Fusion.

This thesis describes in detail an experiment to optimize the weighting of external, internal and ELINT information using the RL E/I Data Fusion Testbed. Internals are RL audio processing functions, such as: Speaker and Language Identification. Externals are other information that can be obtained from the audio signal such as: Transmission Frequency, Direction, and Modulation type. ELINT is the process of observing radar signals to obtain information about their capabilities.

This thesis also describes a series of experiments which can be performed on this Testbed. A preliminary experiment was designed and performed to determine the best set of externals for the Testbed.

14. SUBJECT TERMS
Speech Processing, Data Fusion, Speaker Identification, Language Identification

15. NUMBER OF PAGES
112

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	U/L

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2 89)
Prescribed by ANSI Std. Z39-18
298-102

PREFACE

The Rome Laboratory (RL) External/Internal (E/I) Data Fusion Testbed was a product of an RL research contract with HRB Systems. This preface describes the job of an RL program manager, the requirements for progression of an RL program manager into upper management, and how the courses associated with this MS in Business degree have benefitted the E/I program and this thesis.

Within Rome Laboratory, there are a large number of funded and cooperative research projects with industry or universities. The E/I Data Fusion Testbed contract, the subject of this thesis, is an example.

An RL engineer's job as program manager on these research projects is to perform two functions. These functions are to manage the contract technically and administratively.

Administratively these projects require monitoring of schedule and cost. HRB provided monthly status reports containing a current, detailed list of contractual milestones and actual vs. planned expenditures. Typically, the schedule is provided as a Gantt chart and the funds portion contains expenditures for the next three months in a Government standard Contract Funds Status Report. SUNY Graduate Classes that allowed for a better understanding of these items included Project Management (Meredith and Mantel, 1989) where Gantt charts were covered in detail and Accounting (Weygandt et al, 1990) which dealt with all aspects of direct and indirect costs.

Technically, it is necessary to understand and work as a team with HRB to develop an approach to solve the Government's research and development needs. Project Management allowed for the development of a project which simulated a

manager's performance on a contract from conception to delivery. This activity allowed for a better understanding of the tasks the HRB program manager performs on an RL contract. Other essential skills of conflict resolution when contractual problems arise, effective communication skills to deal with a contractor, and group behavior were dealt with in Organizational Behavior (Robbins, 1991).

Technically, research associated with writing a paper in the Research Seminar (Parker, 1993) and its continuation for this thesis benefitted the E/I Data Fusion Testbed contract. While planning and performing experiments, several things were uncovered about E/I that were not understood and required clarification/explanation by the contractor to RL. These are further described in Chapters 3 and 5 of this thesis. Also, while thinking about how to test E/I (see Chapter 4), these concepts were passed to the contractor before submission of the required test plan. In addition, while documenting the individual components of E/I (see Chapter 2), several pieces were found to be ambiguously stated and not well understood and generated questions that were answered by the contractor. Finally, while planning experiments, different summary statistics were needed. These needs were discussed with the contractor and later added to the Testbed.

Though oriented around business concepts, summary statistics used during this thesis and discussed in the Research Seminar paper were reviewed during a class in Statistics (Moore, 1989). These concepts included mean, standard deviation, median, mode and time series analysis techniques.

Another important RL program manager's job is the technical and cost evaluation of competing contractor responses to an RL request for proposal. This involves a certain level of technical expertise to understand the contractor's proposed approach. Data Communications (Silver and Silver, 1987) enabled a better technical understanding of modems and communications channels, both related to speech processing. In addition, a class in Database Management (Stamper and Price, 1990) allowed a better understanding of a vital part of speech processing since RL is in the

process of creating two unique databases for research and development (Lambert et al, 1993).

Frequently, work breakdown structures, Gantt charts, Pert charts and other management items are also included in a technical proposal. A general organizational chart is often contained along with a summary of business goals of the company. Aspects of the Project Management and Organizational Behavior classes contributed to a better understanding of these portions of the proposal. If available, corporate financial statements could also be examined to determine if the company is stable enough to be awarded the contract. Elements of Managerial Finance (Gitman, 1991) and Accounting would aid in this analysis.

In addition, the proposed costs are evaluated from a technical viewpoint for adequacy and necessity. Though RL has their own financial and cost analysis personnel, a certain high level understanding of Accounting, Financial Management (Anthony et al, 1992) and Managerial Finance provides for a better understanding of direct costs, indirect costs, cost of money, profit, loss, labor rates and categories, overhead costs, travel expenses, etc.

UPPER MANAGEMENT AT RL

Migration from program management to RL upper management has two different tracks: technical and managerial.

Both tracks require a minimum of 12 business credits for Government-wide certification at the correct level for these positions. Classes in Accounting, Finance, Macroeconomics, Microeconomics, Business Law, Marketing, Organizational Behavior, and Statistics count towards this requirement. This degree has provided the classes to fulfill this requirement.

In the technical track, RL personnel desire to achieve the title of group leader or technical director. These individuals strive for world class technical stature in an area, with some small sense of business skills. The computer science electives are not geared

for personnel in this track since they are taught with a business orientation.

In the managerial track, a high-level, non-detailed understanding of technical aspects are required with a maximum of managerial skills. This track involves management of people, not technology. Program and project managers prepare budgets and track financial obligations and expenditures. The budget process was described in Accounting and Financial Management. Supervisors perform performance appraisals, deal with upper management and multiple personalities, make managerial decisions, hire personnel, and approve decisions of program managers. These areas were dealt with in Organizational Behavior.

Finally, in the new President Clinton initiative, developing civilian uses of military technology is extremely important. In addition, the local RL politicians are attempting to employ this strategy to transfer as much RL technology to the local community as possible to minimize the chances of the laboratory moving to other locations. As a result, a background in Marketing (Boone and Kurtz, 1989) strategies and techniques is beneficial for RL engineers.

Accesion For		
NTIS CRA&I	√	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

ACKNOWLEDGMENTS

There are numerous people I wish to thank for their help during the time period this thesis was written. Dr. Roger Cavallo is thanked as my thesis advisor for allowing me to focus this thesis, which at one time was a bunch of words and lists with no common thread. The engineers at HRB Systems Inc., specifically Dave Morgan and Mark Shirey, are thanked for their operational, software, and External/Internal support without which this thesis could not be completed. Norm Lambert of Mei Technology Corp. is thanked for answering my endless operational questions. A thank you goes to my fellow RL workers, specifically John Grieco and Doug Smith for their UNIX expertise and my group leader, Jim Cupples, who in our numerous technical discussions caused me to sit back and think a little more. Lt. Mike Perz from the Air Force Academy is thanked for his assistance in the data collection and analysis for the preliminary experiment. Teresa Casanova typed this thesis and an endless number of experimental data points into Excel spreadsheets, and is graciously thanked. Finally, last but not least, a huge thank you goes out to Sharon Parker who also assisted me with the spreadsheet data entry but mostly stood by me and gave me words of love and encouragement when most needed during this thesis.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter 1

INTRODUCTION

Rome Laboratory (RL) has been performing speech processing research and development since the early sixties. Areas of research include speech recognition, speaker identification, language identification, keyword recognition, platform identification, and external/internal (E/I) data fusion.

Data fusion for this thesis refers to the merging of information produced by three independent data sources: externals, internals, and Electronic Intelligence (ELINT) information. The RL E/I Data Fusion Testbed is a system which allows for experiments to be performed to merge and correlate these data sources. These terms are further defined in Section 1.1.

This objective of this thesis is to perform an experiment examining the weighting of decisions made by the external, internal, and ELINT processes. It examines the relationships between decisions made by these processes and gives some relative importance to these decisions. It was chosen due to its ease of use, and relevance to both the speech processing and operational communities. This experiment is described in detail, including its design, analysis and conclusions.

This thesis also describes the history of the Testbed development and its general approach to data fusion. The thesis also details a preliminary experiment to get more familiar with the capabilities of the Testbed. Finally, the thesis describes many other experiments which could be run with the Testbed.

1.1 DEFINITION OF TERMS

Before describing the E/I Data Fusion Testbed, it is necessary to understand some

of the terms which will be used throughout this thesis. This section provides a definition of these terms (Shirey and Morgan, 1993).

Internals are processes that analyze an audio signal and result in a series of decisions and scores. Externals are characteristics of a voice communication other than the audio signal. Since the intent of this thesis is to perform experiments with the E/I Testbed, the technical details and performance levels of the individual externals and internals are not described in any detail.

1.1.1 INTERNALS

The internal speech processing functions are the RL speaker identification, language identification, platform identification, and keyword recognition systems. Each system makes decisions based on characteristics extracted from the audio signal and is in a different stage of research and development. Details on the state of the art of these processes can be obtained from the yearly Proceedings of the IEEE Acoustics, Speech, and Signal Processing Conference.

Speaker identification is the uncooperative recognition of a speaker based on features extracted from the speech of an audio transmission. This technology is being researched by RL independent of language and the words that are spoken.

Language identification is the uncooperative recognition of the language based on features extracted from the speech of an audio communication. This technology is being researched by RL independent of speaker and the words that are spoken.

Platform identification is the function of identifying the platform of a speaker based on features extracted from the background noise of an audio communication. For our purposes, platform can refer to differentiating a helicopter from an airplane or specifically determining if the platform is an F-16 or a B-52 aircraft.

Keyword recognition is the recognition of words and phrases in a communication. This can be performed in a speaker dependent fashion, where every word to be searched for is spoken at least once by a specific speaker. Speaker independent refers

to the case where a general speech model is formed for each word in the database. The tie-in between E/I and keyword recognition is illustrated in Section 1.2.

1.1.2 EXTERNALS AND ELINT

Table 1 gives a hypothetical example that illustrates the externals. This table represents short news reports on President Clinton obtained from several radio station listeners at various times during the same day.

FREQ.	MOD.	LOC.	TIME	SPEAKER	SHORT SUMMARY
1310	AM	ROME	1100	SPK1	LOCAL NEWS REPORT ON
					CLINTON RE: ECONOMY
1350	AM	NY	1110	RUSH LIMBAUGH	EDITORIAL RE: CLINTON'S
					VIEWS ON HEALTH CARE
103.4	FM	NY	1500	LARRY KING	DISCUSSION ON CLINTON'S
					DEFENSE STRATEGY
105.3	FM	BOS	1600	DAN RATHER	DISCUSSION ON ECONOMY

TABLE 1: EXTERNAL/INTERNAL EXAMPLE

The externals in the Testbed are frequency, modulation type, radio station location, radio type, and direction with respect to the listener. In Table 1, the radio station's frequency (FREQ.), modulation type (MOD.), and location (LOC.) are recorded based on the position of the radio dial where the news report was found (e.g., Rome station WTLB 1310 AM). The other two externals refer to the direction (DIR.) of the station with respect to the listener measured in degrees from true North and radio type (RT. - walkman, stereo, etc.).

External systems are not totally accurate when measuring the frequency and direction parameters. In the E/I Data Fusion Testbed, these externals are used subject to error tolerances. For example, occasionally the radio station located at frequency

103.4 MHz can be heard at 103.5 MHz.

Electronic Intelligence (ELINT) reports are also included in the Testbed. These separate reports contain the type and location of the desired radio stations and are determined by specially designed ELINT systems. The calculation of the ELINT location parameter is also subject to error tolerances.

1.1.3 OTHER DEFINITIONS

To insure less duplication of Table 1 news reports due to multiple listeners scanning the radio at the same time, a routing table could be developed. This table specifically assigns speakers, languages, and news events to listeners for automatic radio station search. For instance, if the routing table for listener one contained the speech of Rush Limbaugh, listener two could skip that broadcast automatically and go on to some other news report. In the E/I Data Fusion Testbed, these tables can be easily created and modified as described in Section 5.3.6 and are the source of the routing statistics described in Section 2.3.

RL technology is being developed for tactical speech processing applications. This means that the decisions made by the speech processing systems must be determined in as little time as possible. Strategic applications have the luxury of longer decision and training times since there are less rigorous time constraints on the results.

Finally, the words activity and scenario must be defined. In Air Force (AF) terms, an activity is an event for which a report is created. A scenario is the sequence of events which are reported. A typical AF scenario is composed of several activities. Table 1 is also called a scenario report and the short summaries are representative of the activity.

1.2 EXTERNAL/INTERNAL MOTIVATION

Waltz and Llinas state that the objective of data fusion "is to derive more information through combining, than is present in any individual element of input data" (Waltz and Llinas, 1990). The concept of E/I data fusion evolves from this definition. This

section uses information extracted from Hamer and Foil (1987) and Shirey and Morgan (1993).

Considering only the internals, interrelationships between speaker identification and keyword recognition can be demonstrated since speaker dependent keyword recognition produces better results than speaker independent recognition. Thus, if the speaker is accurately recognized via speaker identification, then speaker dependent recognition can be performed by loading in the proper speaker's keyword training templates.

Further, people speak a limited number of languages. Thus, if accurate identification of the speaker can be obtained, then these internal-internal interrelationships (contained in an associations table) can be used to verify the results of the language identification system. This illustrates the data fusion relationships between language identification and speaker identification.

To illustrate the concept of externals' providing information on speakers of interest, observe the scenario report in Table 1. In this report, interrelationships are built between externals and speakers. Similarly, these interrelationships could be developed between externals, languages, and platforms. These interrelationships, expressed in an external relations table, provide the required information to perform the correlation of externals with internals. Similarly, ELINT relations tables can be developed to associate ELINT data with internals. The creation and use of these relations tables is described in Section 2.2.

1.3 EXTERNAL/INTERNAL HISTORY

The history of the RL E/I program parallels the USAF programmatic cycle for research and development. This cycle begins with a 6.1 basic research stage in which theoretical research without application is performed. In the 6.2 Exploratory Development stage, an algorithm is developed on some computer for experimentation, test, and evaluation without regard to performance, speed, or any guarantee of success. In the 6.3A stage, an application component is added to the 6.2 stage for this

experimentation, bridging the gap between 6.2 and 6.3 research. In the 6.3B Advanced Development stage, an operational system is designed, developed, implemented, and fabricated around an application for operational test, evaluation, and deployment. The final stage, 6.4, represents Production and Deployment in which several systems are put into the field for operational use.

The rest of this section generally illustrates the chronological history and development stages of the RL sponsored E/I research program. Basic research (6.1) work in data fusion has evolved through other areas of Rome Laboratory, academia, and other government and industrial sponsorship. No sponsorship has come directly from the RL speech processing group.

1.3.1 EXTERNAL/INTERNAL 6.1 PROGRAM

The initial RL 6.1 E/I data fusion program, completed in 1985, performed a study and prepared a report on the data fusion of externals and internals. In the study, (Hamer and Foil, 1987) several externals were identified and defined, and an architecture was developed for E/I data fusion. Figure 1 illustrates this architecture.

Figure 1 shows digitized audio input to the speaker and language recognition systems, each having their own speech based training models. These systems output a series of decisions and scores. In parallel with these internal processes, lists of externals are input to speaker and language ranking and likelihood assignment processes, based around trained external models. Results of these external and internal processes are fused by speaker and language decision logic with an "educated result" produced. The internal-internal interaction between the language decision logic and speaker decision logic is also shown. Following the "educated" speaker and language decision, this information along with the external data is input into the keyword recognition system to select the appropriate word templates. This architecture also calls for the additional

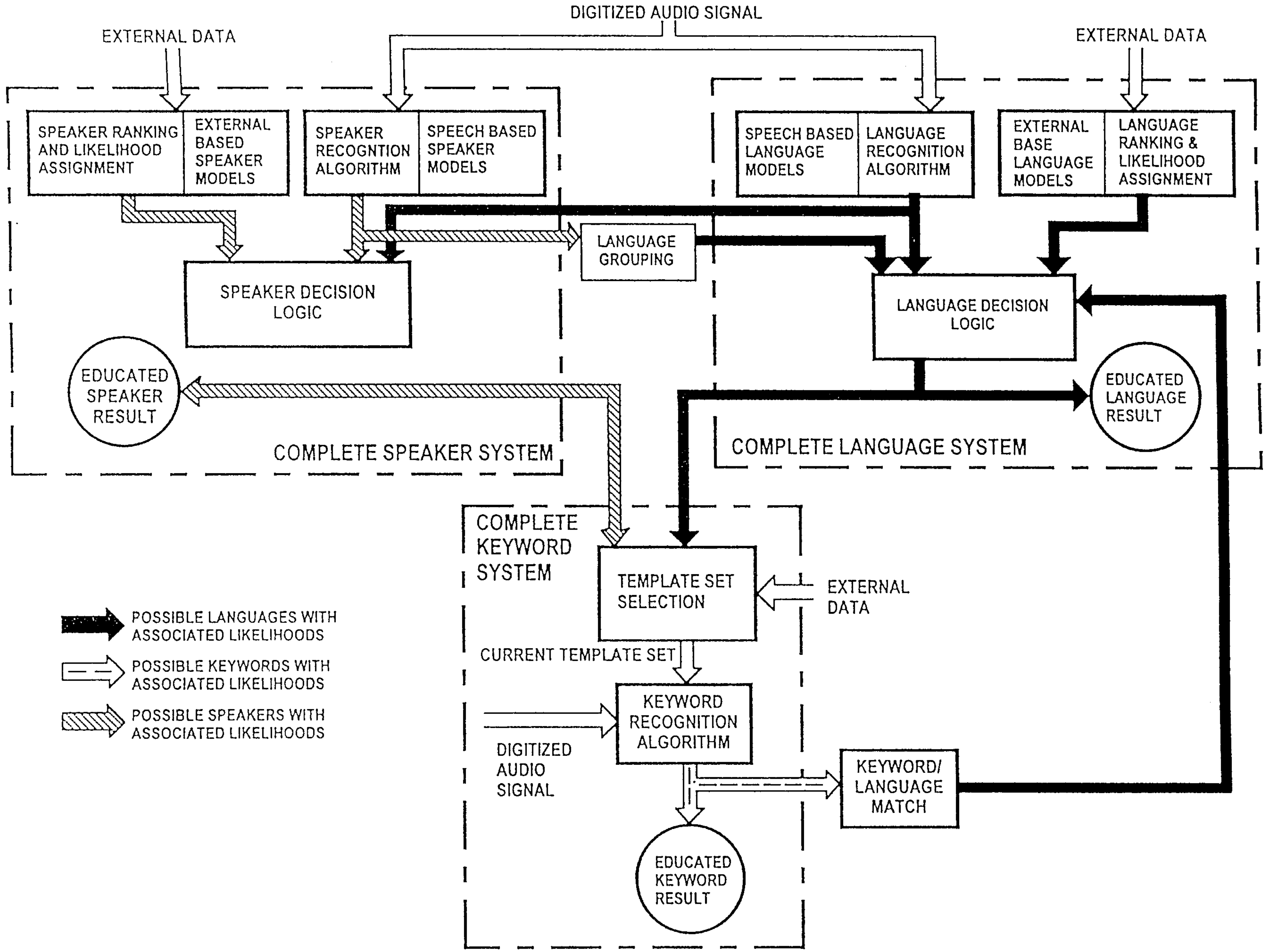

Figure 1: External/Internal Configuration

keyword/language match to insure the language of the keyword agrees with the language identification result. The flow of this architecture is being followed today and is further described in Chapter 2 of this thesis.

1.3.2 EXTERNAL/INTERNAL 6.2 AND 6.3A STAGES

Since the initial study, three Exploratory Development Models (EDM) have been developed. All three have transitioned the 6.1 work previously described into the 6.2 and 6.3A stages.

The initial E/I EDM, (Woodsum and Harms, 1987) completed in 1987, was developed on a VAX 11/780, with a user interface and simulation developed using the Digital Equipment Corporation Forms Management System (FMS).

Two of the deficiencies of this EDM were the inflexibility of this proprietary FMS package and the difficulty to add, modify, and change externals and internals. Three data fusion strategies included were Bayesian, addition of external and internal scores, and addition of external and internal ranks.

To overcome the problems with the first E/I EDM, a second research program was funded and completed in late 1989 (George, 1990). This EDM was developed on a SUN workstation, with Inference Corporation's Automated Reasoning Tool (ART) expert system shell performing rule based data fusion.

An operational database was created for realistic simulation. This database was created from twelve AF activities which were organized into nine different scenarios. All scenarios have a complete record of the conversations that occurred, with the speakers, languages, and externals accurately denoted and stored as ASCII scenario ground truth files. These files are the source of the simulated inputs described in Section 2.1 of this thesis. Finally, audio tapes were produced which could be input into the actual speaker identification and language identification systems to create live internal results rather than the simulations.

A simulation was created of the external, speaker identification and language

identification functions on an INTEL 310AP microprocessor. This simulation was created to allow the experimenter to select and vary score sequences of externals, speakers, and languages in a fashion similar to the actual external and internal systems. The simulation allowed for the selection of scenario groups (1,2,3), (4,5,6), or (7,8,9). Following selection of the scenario group, all correct speakers and languages known by the software were assigned a confidence score of 90. All other speakers and languages in the scenarios were assigned an identical confidence score of 50. The experimenter could then tediously go through one by one decreasing the correct internal score and/or increasing the incorrect ones to simulate errors made by the internals. The alternative to this tedious process, is to enter numerical values for distortion which randomly adds or subtracts an amount from each score in a fashion controlled by the software in the INTEL. This concept is the basis for the simulation which is described in Section 2.1 of this thesis. Externals could be altered in the same manner.

Fusion was performed using a rule based expert system, coded using the LISP based ART shell. The only internal analyzed was speaker identification since no rules were written for language identification. In this system, the data fusion algorithm checked the magnitude of the highest ranking speaker identification confidence score. If this score was higher than a defined threshold, it was reported without performing any fusion. If the score was less than the threshold, then the system performed rule based data fusion.

The only external used in this data fusion process was the location of the speaker measured in terms of latitude and longitude. The rules checked to see which speaker was closest in distance to the simulated location. The system kept track of the average location of each speaker in the database and also measured the variance. If the variance was above a threshold, the speaker was designated as moving and eliminated from consideration.

The algorithm improperly assumed that since the top speaker score was less than the threshold, it was not correct. As a result, the computer software automatically

eliminated the top scoring speaker from consideration by the data fusion algorithm.

There were two major difficulties with this data fusion algorithm. First, the top speaker was automatically accepted when its score exceeded the threshold. What if that speaker was not likely to be at the location? Second, the logic to automatically eliminate the top scoring speaker from consideration was flawed. What if the speaker was closest in distance to the reported location? Overall comparison of speaker identification performance before data fusion versus after resulted in a decrease in performance due to these difficulties.

Another flaw of this EDM was that all software was written using the LISP based ART shell. This included the data fusion algorithm, the displays, the record keeping routines and the SUN communication protocols with the INTEL. As a result, the system collected an enormous amount of LISP generated "garbage" as it ran. After less than twenty minutes of processing, the system stopped for "garbage collection". This characteristic made the system useless.

Finally, the EDM only considered the speaker internal and the location external in its data fusion process. A more elaborate rule based expert system would be required to make this system operationally useful. These deficiencies were corrected in the third E/I EDM.

Chapter 2

EXTERNAL/INTERNAL DATA FUSION TESTBED

This chapter explains in detail the current RL E/I Data Fusion Testbed, the third E/I EDM (Shirey and Morgan, 1993). When new terms are defined, they are underlined, as in Chapter 1.

Figure 2 is provided as a high level design for creating and running experiments on the Testbed. This high level design begins with the selection of scenario and simulation parameters. The External/Internal/ELINT simulation algorithm then turns the inputs into a sequence of LSAP (language, speaker, activity, and platform) decisions. The relations table, which contain the external-internal and ELINT-internal interrelationships, are input to their respective Processor algorithm to obtain LSAP decisions based on these inputs. Finally, both lists of LSAP decisions are merged and correlated by the E/I Data Fusion algorithm to produce LSAP decisions which are stored in the system audit file along with other intermediate outputs illustrated in Figure 2. Another input to the Data Fusion algorithm is the internal-internal interrelationships in the associations table. Each of these parts of the Testbed are described in detail in the sections which follow.

Experiments on this Testbed can be conducted to change the inputs (outside the boxes in Figure 2) and determine their effect on the LSAP decisions recorded in the system audit file. For example, an experimenter can select different scenarios and different values for the E/I simulations and evaluate statistics which can be obtained from results in the audit file. A separate experiment could process different inputs through the four data fusion algorithms to assess the performance of these algorithms.

This Testbed has the flexibility to run a number of different experiments. A short

experiment is described in Chapter 3 of this thesis. A summary of ten possible experiments is given in Chapter 4, and one particular experiment is designed and performed in Chapter 5.

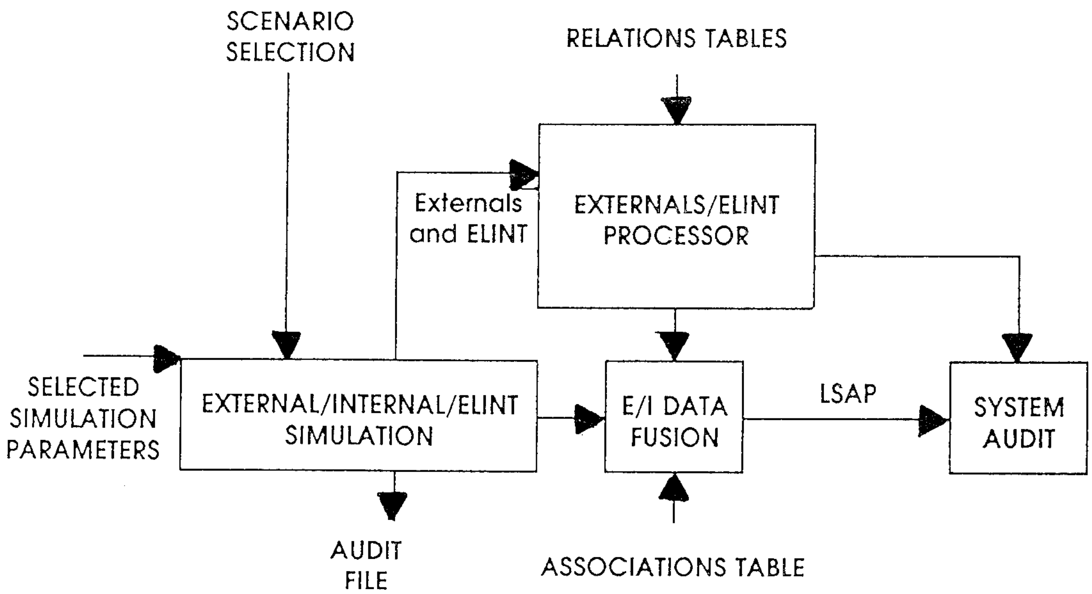

FIGURE 2: EXTERNAL/INTERNAL TESTBED COMPONENTS

2.1 EXTERNAL/INTERNAL/ELINT SIMULATION

In the Testbed, the same nine scenarios and ground truth files as described in Section 1.3.2 are included. Ground truth files are the correct external and internal values identified from the created scenarios. Platform identification and ELINT values were added to these ground truth files in the third EDM.

The ground truth files are an integral part of the External/ELINT/Internal simulation algorithm. They are also used to create the associations table, external and ELINT relations tables and as the correct answers in the creation of statistics in the audit trail.

The experimenter initially selects any number of these nine scenarios. For example, in the experiment in Chapter 5 of this thesis, three groups of three scenarios were used.

In addition to scenario selection, the experimenter must select the simulation

length. If zero minutes is selected for the beginning time and five minutes is selected for the end time. the simulator will only generate the first five minutes of the selected scenarios. Each scenario is approximately 45 minutes in length.

Another parameter of the simulation to be selected by the experimenter is the decision report time. This parameter controls the maximum length of time before the Testbed makes a final decision. This parameter is adjustable from one to five seconds. These times are based on the average transmission length in the tactical speech processing environment. If this parameter is set for five seconds and a transmission is only two seconds long, an E/I LSAP output report is created without waiting for five seconds to elapse.

Another variable to be selected in the simulation is one of the four flight paths named A, B, C, and D. The only parameter value changed as a result of this selection is the direction external.

The experimenter can select any combination of externals and/or ELINT desired for an experiment. The five externals as described in Chapter 1 are location, direction, frequency, radio type, and modulation type. In addition, different presets are provided to automatically choose popular external/ELINT combinations.

A confidence score can be assigned for each external/ELINT simulation. This score represents the experimenter confidence in the measurements provided by the external and ELINT simulation.

2.1.1 INTERNAL SIMULATION

Separate simulations are provided for the speaker identification, language identification, and platform identification internals; however, the functionality of each simulation is identical. Each internal simulation allows for the selection of the number of possible internals (e.g., 20 languages, 300 speakers, 20 platforms). A series of internals and confidence scores is created. The maximum numbers for these simulations were chosen to be operationally significant.

A second parameter allows for the experimenter to choose the number of confidence scores to be input into the data fusion algorithm. Thus, if the 300 speaker simulation only has 50 valid scores, there is no reason to provide all 300 for fusion.

The internal simulation algorithm can produce constant ground truth and a series of constant secondary scores. For instance, if there are four languages in the simulation and a score of 80 is assigned to the ground truth language (L1), and a 50 to the rest, the simulator output will be L1 80, L2 50, L3 50, and L4 50. All other language score sequences will be produced with the same identical pattern. This simulation is called a constant simulation.

A constant simulation allows for the controlled simulation of internals since the score patterns are exactly identical and the ground truth internal can be assigned a value higher than the rest creating a perfect simulation. Though this method of score generation is good for debugging algorithms, it is operationally unrealistic. Normally, the internal confidence scores are not constant and in the tactical speech processing environment the internals err.

Thus, the internal simulation allows for the selection and addition of distortion. Distortion allows for a random number to be added or subtracted to each of the secondary scores without any alteration of the ground truth score assigned to the correct internal. For example, selecting a distortion of 31 would produce the above list with a random number between 1 and 31 added or subtracted from the secondary scores: L2 81, L1 80, L3 40, and L4 30. As illustrated, using this method the internals simulate errors in their list since L2 now has a higher value than ground truth language L1 .

Another way of simulating errors in the internals is by selecting a random simulation. In this type of simulation, the secondary scores would be crested by a random number generator. Assuming L1 is the ground truth language with the score assigned as above, a report such as this could be obtained: L2 95, L1 80, L3 3, and L4 1. Distortion could be selected by the experimenter to further alter these scores.

No matter whether constant or random is chosen for the simulation, the generation of internals creates choices around ground truth. If L5 is the ground truth language and three languages are chosen, the simulation produces scores for L4, L5, and L6. This definition assumes that the languages that are next to each other in the database are typically confused with each other.

The final important characteristic of the internal simulation is the ability to reproduce identical internal reports from experiment to experiment by the selection of reproducible rather than dynamic scoring. If an experimenter selects dynamic scoring, a different scoring pattern is used for each experiment. The selection of reproducible continuously produces the some score pattern.

2.2 EXTERNALS/ELINT PROCESSOR

As depicted in Figure 2, the externals' values and confidence scores are input into the Externals Processor. Similarly, ELINT data is input into the ELINT Processor. Another major input is their respective relations table which is established by experienced radio station listeners in a training session.

The process of creating ELINT and external relations tables is identical. These relations tables are established as a matrix between all possible external or ELINT values and internal values. For example, if the experimenter is interested in how frequency contributes to the recognition of Rush Limbaugh, all radio stations of interest across the country must be searched and the matrix filled in upon detection of his voice on a certain radio frequency.

The scenario report in Table 1 illustrates the relationships between speaker identification, frequency, modulation type, and location. These relationships can be recorded over multiple days and used as the basis for the relations table data. Similar relations can be developed for the other internals, externals, and ELINT.

Following the accumulation of the raw data, normalization of the numbers occurs to produce weights between one and four. An experienced listener familiar with the

association between Rush Limbaugh and radio frequency can look at these final weights and adjust them based on this knowledge. This adjustment is to only be used in extenuating circumstances. For instance, if we know Rush Limbaugh is **always** on radio frequency 1310 at 12:00 noon, the expert can provide a weight of 100 in the relations table for this interrelationship.

The Processor uses these relations table weights, preestablished relative weights, and the simulated external values with confidence scores to mathematically obtain a total score for each internal in the matrix. All scores are then normalized between 0 and 100. The final output of the algorithm is an external based ranked list of Language, Speaker and Platform Identification (LSP) choices and confidence scores. In addition, the Externals Processor makes an assessment of the activity (A) from the external data presented, creating external based LSAP decisions and scores when linked with the LSP.

A separate ELINT Processor produces an independent set of LSAP decisions based on ELINT data. ELINT reports occur at different times than the simulated external and internal reports. In order to perform the LSAP correlation with ELINT at the same time as the Externals Processor, an initial check is made for relevant ELINT data. If an ELINT report occurs within a time threshold and within a distance threshold of the simulated external and internal report, then the ELINT data is relevant. The mathematics and normalization of this relevant data into ELINT based LSAP decisions and scores is identical to the Externals Processor.

2.3 EXTERNAL/INTERNAL DATA FUSION

Among the inputs to the E/I data fusion algorithm are the external and ELINT based LSAP decisions and scores as described in Section 2.2. Other inputs are the internal LSP score sequences provided by the simulations described in Section 2.1. The final input into the data fusion algorithm is an associations table developed by accumulation of the known internal-internal interrelationships during a training session by an experienced radio station listener.

A file named eidefaults contains default parameters and filenames for the data fusion algorithm. These parameters are listed in Appendix A with the system default values set by the contractor (Shirey and Morgan, 1993).

Four data fusion and correlation algorithms were developed. Each algorithm can be independently tested and analyzed by the experimenter for performance differences.

Each data fusion algorithm performs a merge of the three input lists of decisions and scores. Each list is weighted independently from 0 to 100. These weights are contained in the eidefaults file and can be altered by the experimenter. No theoretical work or experiments have been done to determine optimum values for these weights. The experiment in Chapter 5 of this thesis analyzes these weights in detail.

For example, to merge language identification (LID) results, W1 represents the amount of weight provided to the results of the internal LID simulation, W2 represents the amount of weight provided to the external based LID results, and W3 represents the amount of weight provided to the ELINT based LID results. If A1 represents the score for English provided by the LID simulation, and B1 represents the score for English provided by the external based LID component, and C1 represents the score for English provided by the ELINT based LID component, the score provided by the merge is calculated by the following formula:

$$\frac{W1*(A1)+W2*(B1)+W3*(C1)}{W1+W2+W3}$$

This merge takes place for each language in each of the three lists. In addition, this merge occurs in the same manner for speaker and platform identification. For activity identification, since there are only external and ELINT based inputs, W1 and A1 are zero. When ELINT information is not used, C1 is zero and W3 is its default weight.

This merge algorithm is one of a larger class of algorithms which could have taken the three lists and produced one output sequence. The score sequences could have

been multiplied together in probabilistic Bayesian sense, combined using radio station inference rules, or processed using a blackboard expert system but the RL contractor evaluated different schemes and chose to implement the merge previously described due to its computational simplicity.

The initial data fusion algorithm, Correlator 1, merges the input list of decisions and scores by equally weighting all decisions (W1, W2, and W3=100). Following this initial ranking, all combinations of the top "beta" (set in the eidefaults file) number of LSAP's are mathematically scored. The final step discards any resulting LSAP combinations that do not appear in the associations table.

The second data fusion algorithm, Correlator 2, added a capability to change the weights of the merge to any number between 0 and 100. For instance, if the externals provide a poor measure of language identification, the external based LID scores could be given a small weight to de-emphasize them in the merge algorithm.

Following the selection of the top "beta" number of LSAP's as in Correlator 1, all resulting LSAP combinations are generated and scored. The algorithm then eliminates one item in all remaining LSAP combination and scores it with this item removed. For instance, if the combination (L1, S1, A1, P1) was generated, the algorithm would score (L1, S1, A1, *), (L1, S1, *, R1), (L1, *, A1, P1), and (*, S1, A1, P1). A later version of the correlator will attempt to fill in the "*" to recover a correct hypothesis not generated by the algorithm. For example, (*, S1, A1, P1) would be scored as (L5, S1, A1, P1), with the "*" replaced by the correct language L5.

In addition, Correlator 2 keeps **all** good scoring LSAP combinations even if they are not in the associations table but discards hypotheses which are not within "cutoff" (set in the eidefaults file) percent of the top scoring hypothesis. For example, if the top scoring LSAP had a score of 60 and the "cutoff" was 5% then any final hypothesis with a score of 56 or less would be discarded.

The third version, Correlator 3 performs the merge and generates hypotheses as Correlator 2 but it assigns the missing item ("*" above) a numerical score. In addition,

the algorithm generates combinations with two missing items as above and attempts to find good scoring hypotheses. A new parameter, "gamma", (set in the eidefaults file) is used to check the generated hypotheses against the associations table entries and discard any invalid combinations. Following this step, an iterative algorithm attempts to fill in all LSAP missing items as described above. The final step keeps all scoring hypotheses within "cutoff" percent of the top scoring hypothesis.

Correlator 4 implements the same algorithm as Correlator 3 but generates hypotheses in a different manner. Correlator 3 performs the merge, then hypothesis generation. Correlator 4 performs hypothesis generation on the outputs of the Externals Processor, ELINT Processor and internals then merges the generated hypotheses. If the hypothesis is on all lists, the resulting scores are averaged. If a hypothesis from one list matches a hypothesis on the other lists with a missing item, put the complete hypothesis on the final hypothesis list with the scores averaged. Finally, if the hypothesis was only on one list, the scores were averaged with zeros. Only hypotheses within "cutoff" percent of the top scoring hypothesis are reported.

In all four Correlator algorithms, multiple hypotheses result when answers are within "cutoff" percent of the highest scoring hypothesis. The highest scoring hypothesis is displayed on the Testbed output screen and all final LSAP hypotheses are recorded in the system audit file.

2.3 SYSTEM AUDIT FILE

For every experiment performed on the Testbed, an audit file is created. In this audit file, all simulation, ELINT Processor, Externals Processor and correlator initialization parameters are recorded along with the routing table. For every transmission, all simulated internal, ELINT, and external inputs are recorded, as well as the outputs of the Externals and ELINT Processors. Also included are the Testbed final LSAP hypotheses and the scenario ground truth. All outputs which do not agree with the scenario ground truth are denoted so the researcher can try to determine why the error occurred.

The end of the system audit file has some summary statistics for the experiment. All internal summary statistics represent the percentage correct internal identification provided by the simulation. The ELINT and Externals Processor statistics indicate the percentage accuracy of the internals given only ELINT and external data. In addition, the Testbed provides merge algorithm summary statistics describing the accuracy of the top scoring LSAP combination after the merging of all inputs. Other correlation statistics after hypothesis generation provide the accuracy of the internals contained in the final best scoring hypothesis and statistics on the accuracy of the final LSAP hypotheses. Finally, routing statistics are provided to show how accurate and efficient the routing decisions are made by the E/I Data Fusion Testbed. The statistics in this audit file are invaluable to the experiments described in Chapters 3, 4, and 5 of this thesis.

2.4 MISCELLANEOUS

The Testbed allows for the choice of speed. In a demonstration, normal speed will play the scenarios out at the realistic scenario time. For experimentation, speeds of 2X, 4X, and maximum (Max) are provided. The 2X and 4X cases are clocked to provide that much speed up to the scenarios. In the Max case, the simulated input records are provided to the correlator as fast as possible, resulting in three different scenarios (each composed of 45 minutes of data) being processed in approximately 7 minutes.

Chapter 3

INTERIM EXPERIMENT

This chapter describes the design, performance and conclusions drawn from a quick experiment performed with the Testbed. This test was designed and performed to demonstrate that the Testbed could be used in experimental analysis. The experimental conclusions described are not proven, just measured as trends since the amount of data processed in this quick test is not significant. Lessons learned from this quick experiment are followed in the thesis experiment described in Chapter 5.

3.1 INTERIM EXPERIMENT OBJECTIVE

This short experiment was designed to determine the effect of **all** possible external combinations on internals using data from three different scenario groups. In this experiment, the number of externals were varied in combinations of 5, 4, 3, 2, 1, and 0. The results were expected to vary as the scenarios and external combinations varied, but it was hoped to find trends which would indicate the value and reliability of the externals and their thirty-two different combinations.

3.2 INTERIM EXPERIMENT DESIGN

During experimental design, each of the internal and external simulations were required to be set up to allow additional familiarity with the various simulator internal and external input options. In addition, before running this experiment additional familiarity was gained on what the different audit file summary statistics meant.

All audit file summary statistics were examined for applicability to this experiment. All routing statistics are dependent on the routing table. Since this experiment was meant to be quick, there was no time spent developing a realistic routing table, thus

these statistics were not used. For the experiment in Chapter 5 of this thesis, a realistic routing table is created.

The merge algorithm summary statistics were not examined for applicability. In the design of the experiment in Chapter 5, these statistics are considered.

In this quick experiment, the ELINT and External Processor statistics illustrating the accuracy of the Externals Processor and ELINT Processor were not as interesting as the difference between the simulation LSP results provided in the internal summary statistics and the final LSAP results in the correlation statistics. The difference between these results measures the amount of improvement gained by using the correlation algorithm. This difference was used extensively in this quick study and will be used in the experiment in Chapter 5 as well.

It was also decided to calculate the average improvement over three different scenario groups for comparison purposes. In the long run, these statistics are more reliable than basing them on only one single group. Statistical means and multiple iterations are also used in the experiment in Chapter 5.

3.2.1 INSURING TESTBED CAPABILITIES

During experimental design, it was deemed essential to perform a quick test to insure that several of the system capabilities were properly functioning. These tests were not extensibly performed.

By making multiple runs using the same parameters, reproducible scoring for the simulations was proven to provide the same data repeatedly by looking at multiple tests with identical summary statistics. In addition. when exiting and reentering the Testbed, these statistics still were the same. This reproducibility was also tested by saving the experiment and looking at the summary results after later recall. Though reproducible worked for this experiment, Chapter 5 will illustrate that reproducible tests can run only in this one specific case.

Constant ground truth and secondary scoring was verified by setting the

language and platform identification simulations to constant scoring and observing the nonvarying scores in the audit file report.

Final audit file statistics were proven correct in a ten transmission run. Each statistical category was analyzed by hand and compared with the audit file results.

3.2.2 INTERIM EXPERIMENT SETUP

The test was designed to create a speaker identification accuracy between 50 and 75% and measure the improvement due to the correlation algorithm. A reproducible simulation of twenty speakers was chosen, with the ground truth score of 90. Random scoring was selected for the other nineteen speakers with no distortion added. Only five scores were reported.

The language and platform identification simulations were chosen as constant ten item simulations with no distortion (100% accurate) to minimize the number of varying conditions. The externals were chosen for each experiment in accordance with the objective and simulated with 100% confidence. No parameters in the eidefaults file were altered.

Three groups of scenarios were run through all combinations of externals: (7,8,9), (4,5,6), and (1,2,3). The experiments were performed at 2X speed and processed 15 minutes of scenario data. The decision time was set at 3 seconds. The ELINT simulation was not used.

3.3 INTERIM EXPERIMENT RESULTS AND ANALYSIS

Measurements were made of the speaker identification and LSAP accuracy before and after correlation. A sample of the performance results for three different experimental conditions is included in Appendix B. The detailed performance statistics are included in a report by Perz and Parker, (1993). All analysis and conclusions are derived exclusively from the statistics in that report.

Appendix B illustrates the statistics gathered and analyzed. The table shows the speaker and LSP simulator statistics and the subtraction of these values from their

respective correlator statistics resulting in percentage improvements. For example, the LSP correlated score for 5 externals and the scenario combination (1,2,3) equals 83.27%. This value minus the LSP internals correct score (64.94%) results in an 18.33% LSP improvement. Results are also provided for the (4,5,6) and (7,8,9) scenario groups. Provided for each different set of externals are the mean and standard deviation values over the three scenario groups. Other statistics are included in the Appendix that were not analyzed, namely, the average number of hypotheses and routing statistics.

Analysis for the best external combination began by finding the best individual percentage improvement scores for each of the three groups of scenarios. Many inconsistent external combinations were eliminated and Table 2 was derived. Inconsistent external combinations are those values which scored well for one scenario group, but scored poorly in another. For example, the combination modulation type, radio type, and location showed improvement scores greater that 20% for scenario group (1,2,3), but had negative improvements for the scenario group (4,5,6). Seven combinations stood out from the rest in that they had better speaker and LSP improvement scores for all three scenario groups. Mean and standard deviation scores were added to these tables to better compare them.

Both parts of Table 2 show that the external combination of frequency, radio type, and location has the largest mean improvement percentage with very close to the best standard deviation. The value in the (4,5,6) column in both tables is close to the highest while the (1,2,3) and (7,8,9) values exceed all other values in their respective columns.

All combinations in Table 2 include the frequency external (FREQ). The detailed results also show that frequency is the only external which, when used exclusively, results in performance improvements. Further analysis of the scenario ground truth files discovered that no scenario includes a change of frequency which possibly explains this phenomenon.

EXTERNAL COMBINATIONS WITH BEST PERCENT SPEAKER IMPROVEMENTS					
COMBINATION	(1,2,3)	(4,5,6)	(7,8,9)	MEAN	STANDARD DEVIATION
FREQ/RT/LOC	23.51	26.67	16.66	22.28	5.12
FREQ/MOD/LOC	21.91	26.67	16.66	21.75	5.01
FREQ/MOD/RT/LOC	21.12	27.22	16.05	21.46	5.59
FREQ/RT	21.91	27.78	14.19	21.29	6.82
FREQ/MOD	21.51	25.56	14.81	20.63	5.43
FREQ/MOD/RT	21.91	26.11	12.34	20.12	7.06
FREQ/DIR/RT/LOC	21.51	22.78	13.58	19.29	4.99

EXTERNAL COMBINATIONS WITH BEST PERCENT LSP IMPROVEMENTS					
COMBINATION	(1,2,3)	(4,5,6)	(7,8,9)	MEAN	STANDARD DEVIATION
FREQ/RT/LOC	22.71	26.67	16.66	22.01	5.07
FREQ/MOD/LOC	20.72	26.67	16.66	21.35	5.03
FREQ/MOD/RT/LOC	20.32	27.22	16.05	21.20	5.64
FREQ/RT	20.72	27.78	14.19	20.90	6.80
FREQ/MOD	20.72	25.56	14.81	20.36	5.26
FREQ/MOD/RT	20.72	26.11	12.34	19.72	6.94
FREQ/DIR/RT/LOC	20.32	22.78	13.58	18.89	4.76

TABLE 2: PRELIMINARY EXPERIMENT RESULTS

For this database, the lack of frequency change was designed to be as realistic as possible for the scenarios generated. However, typically, one radio station listener monitors multiple frequencies at the same time.

By observing these statistics, it was noticed that the addition of directional information seemed to hinder performance. Analysis showed:

1. With any one other external, many negative improvement scores resulted.

2. In the test with two other externals, only the one with frequency and radio type failed to result in negative improvement scores.
3. Using all externals except for direction resulted in the best improvement scores for a four external combination.
4. Direction plays a role in all negative improvements at one time or another, except for the MOD/RT/LOC and MOD/RT combinations.

Note that both exceptions included in item 4 in the previous paragraph include the modulation and radio type externals. These two externals seem to work rather weakly together. FREQ/MOD/RT and four other external combinations including both modulation and radio types avoid negative improvements because of the presence of the frequency external or the other external information to compensate for deficiencies in the MOD/RT interactions. *Omission* of modulation type for the four external tests results in improvement scores second only to the omission of the direction external. Using only modulation type resulted in the highest negative improvement scores for a single external.

In order to analyze the general theory of data fusion, means and standard deviations for all numbers of externals were calculated as shown in Table 3. Note that the mean speaker improvement scores are slightly higher than the mean LSP improvement scores. In addition, the mean percentage improvement increases as the number of externals are increased.

	SPEAKER % IMPROVEMENT		LSP % IMPROVEMENT	
NUMBER OF	MEAN	STANDARD	MEAN	STANDARD
5	17.96	5.51	17.42	5.31
4	17.35	6.17	16.00	5.26
3	13.14	8.92	9.19	11.56
2	12.01	6.34	6.08	11.28
1	-6.32	8.98	-12.75	11.04
0	-33.45	16.53	-44.80	9.45

TABLE 3: PRELIMINARY EXPERIMENT DATA FUSION RESULTS

3.4 INTERIM EXPERIMENT CONCLUSIONS

Given the poor performance of the direction external with other external combinations, it should not be used in those combinations. Even though the addition of directional data should, according to data fusion theory, provide for better improvement scores, these preliminary experiments question its use. This consideration will be taken into account in the experiment in Chapter 5.

The database and results illustrate performance of the Testbed against scenarios on fixed frequencies. As of yet, there have been no experiments performed to test how the Testbed correlation algorithms react to changes in frequencies. This would involve modifying the database as discussed in Section 4.1.8.

Based on the date in Table 2, the external combination FREQ/RT/LOC is the best combination. Mean speaker and LSP improvement percentages are better than those of all others. This external combination is used in the experiment in Chapter 5.

The modulation and radio type externals can be useful if used in conjunction with other externals as well, but certainly should not be used alone or together by themselves. This recommendation is taken into account in the experiment in Chapter 5.

Based on the results in Table 3, data fusion theory does hold. As the number of externals increased, the better the results got. The difference between four and five externals was not very significant as compared with the others possibly due to the clashes between modulation/radio type and location/direction.

More tests should be run for statistically significant conclusions. This experiment only considered three scenario groups covering all of the possible external combinations, therefore the results obtained cannot compare with the reliability of many tests. In addition, only 15 minutes of data was run as opposed to the full 45 minute scenarios. The thesis experiment in Chapter 5 corrects for these deficiencies.

Lastly, this experiment did not consider the routing statistics or the merge algorithm's summary statistics. The experiment in Chapter 5 defines a pertinent routing table and analyzes the routing statistics as well as the results of the merge algorithm.

Chapter 4

POSSIBLE EXPERIMENTS

This chapter describes multiple experiments that could be performed with the Testbed. These experiments were created exclusively for this thesis. Some ideas were taken from Shirey and Morgan (1993) and Parker (1993).

There is no long term goal for all of these experiments. Some experiments were defined to evaluate different parameter settings of the Testbed. Others were to clarify or quantify unsubstantiated claims made by the contractor. Some experiments were designed to provide some feedback on externals to the operational community. Finally, some were described to provide some feedback to the speech processing scientific community.

Each experimental description includes an initial statement describing the objective, a general experimental procedure including some of the variables chosen to be constant and some of the variables to be varied, and some statement of the expected final result. There was no attempt to define the specific experimental design and performance measurements for each variable (i.e., appropriate values for "cutoff", "beta", and "gamma" were not defined). Finally, at the end of this Chapter a justification is given for which experiment is to be defined in detail and performed in Chapter 5.

4.1 EXPERIMENTS

Ten different experiments were defined in the sections which follow. All experiments will include several three-scenario groups, include the full 45 minutes of ground truth data, and be performed at Max speed. The 45 minute simulation length is

the maximum run length for the scenarios and was chosen to give enough data to provide the maximum amount of scenario variability. Max speed allows for this amount of data to be run in about 7 minutes, allowing multiple iterations to be processed quickly.

The external set, when not explicitly varied in the description, agrees with the best external set chosen during the quick experiment in Chapter 3 (frequency, radio type, location). No additional thought was given as to whether another choice was better.

In the experiments where the correlation algorithm number is not explicitly varied, Correlator 3 was chosen. This correlation algorithm was used the most during RL's development contract with HRB Systems.

When not explicitly varied in the experiments, realistic values will be selected and kept constant throughout all iterations of the experiment for the following variables: correlation merge weights; correlation parameters "cutoff", "beta", and "gamma"; Externals Processor and ELINT Processor relative weights; frequency and location error tolerances; and ELINT time and distance thresholds. These parameter values will be operationally realistic where possible.

For all experiments, the decision report time will be five seconds, as this parameter never changes the results provided. In addition, unless explicitly changed in the experiments, training for the relations tables and the association table will be based on 100% of the database. Finally, the flight path value will be set and not altered unless explicitly changed in the experiment since the direction external was not a part of the selected external set.

In all cases, an appropriate routing table will be defined and used. This routing table will be created in an operationally realistic manner.

The following ten sections describe the different experiments:

4.1.1 CORRELATOR 3 PERFORMANCE

The objective of this experiment is to perform detailed test and evaluation of

Correlator 3. There has been no long term evaluation of Correlator 3 in terms of performance.

General Experimental Methodology: Run several different internal simulations with various numbers of speakers, languages, and platforms. Run some experiments with ELINT. Evaluate performance of Correlator 3 algorithm.

Final Result: Correlator 3 algorithm improves/degrades speaker, language, activity and platform identification accuracy by X%.

4.1.2 CORRELATOR 3 VERSUS CORRELATOR 4

The objective of this experiment is to compare Correlator 3 with Correlator 4. This would show the differences between HRB's correlation algorithm (3) and HRB's version of RL's correlation algorithm (4).

General Experimental Methodology: Run several different internal simulations with various numbers of speakers, languages, and platforms. Run some experiments with ELINT. Run all simulations through both Correlator 3 and Correlator 4 and compare the results.

Final Result: Correlator 3 or 4 is the better correlation algorithm.

4.1.3 BEST EXTERNAL SET

The objective of this experiment is to determine what set of externals gives the most correlation improvement to speaker identification. This is similar to the experiment in Chapter 3 except the ELINT parameter would be added and other adjustments made to make the experiment more realistic. For all iterations of this experiment, platform and language identification would be kept constant.

General Experimental Methodology: Run all combinations of 1-5 externals and different flight paths. Run several different speaker identification simulations with different numbers of speakers. Ideally, run the same simulations through all combinations of externals and compare the results.

Final Result: This combination of externals provides the best E/I correlation

performance.

4.1.4 OPTIMAL PARAMETER SET

The objective of this experiment is to determine what parameters provide the most correlation improvement.

General Experimental Methodology: Run several different internal combinations with a set number of speakers, languages and platforms through different error tolerance for frequency and location; different external, internal and ELINT merge weights; different values for the correlation parameters "cutoff", "beta", and "gamma"; different Externals and ELINT Processor relative weights; and different ELINT relevant time and distance thresholds. Run some experiments with ELINT. Compare the results over multiple iterations.

Final Result: These parameter values provide the best Correlator 3 performance.

4.1.5 LIMITED TRAINING DATA

The objective of this experiment is to determine the effect of limited training for the ELINT and Externals Processor relations tables and associations table on Correlator 3 performance. Some limited training data tables have already been created for this test.

General Experimental Methodology: Run several different internal combinations with a set number of speakers, languages and platforms. Run some experiments with ELINT. Ideally, run the same simulations but vary the amount of data the ELINT and Externals Processor relations tables and associations table are trained on. Evaluate differences in performance.

Final Result: This correlation algorithm's performance improves/degrades by X% when training the tables with Y% of the database.

4.1.6 FOUR CORRELATION ALGORITHM COMPARISON

The objective of this experiment is to compare the performance of Correlator 1 vs. Correlator 2 vs. Correlator 3 vs. Correlator 4. This would show the differences between all four correlation algorithms and conditions for which the algorithms

performed excellently.

General Experimental Methodology: Run several different internal simulations with various numbers of speakers, languages, and platforms. Run some experiments with ELINT. Ideally, run the same data through all four algorithms and compare the results.

Final Result: This correlation algorithm performs well in these circumstances.

4.1.7 LARGE POPULATION SPEAKER IDENTIFICATION

The objective of this experiment is to determine the improvement in large population speaker identification (300 speakers) using E/I correlation. This would also involve developing multiple ground truth files with 300 speakers and appropriate association and relations tables. In addition, this would involve modifying the software to vary the scores of the 300 speaker id's like the current simulation software.

General Experimental Methodology: Run multiple simulations with 300 speakers and constant language and platform simulations. Run some experiments with ELINT. Compare the performance before and after correlation.

Final Result: E/I correlation improves/degrades large population speaker identification accuracy by X%.

4.1.8 NEW EXTERNAL/INTERNAL DATABASE

The objective of this experiment is to determine performance of Correlator 3 on another E/I database. The Testbed is currently built around a specific database with specific ground truth files. This will convert and test E/I against this new database. This will require the development of the following new items: scenarios, ground truth files, associations table and Externals and ELINT Processor relations tables.

General Experimental Methodology: For the new database, run several different internal simulations with various numbers of speakers, languages, and platforms. Run some experiments with ELINT. Evaluate the performance of Correlator 3.

Final Result: The E/I correlation and merge algorithms perform this well on this new database. This demonstrates the applicability of this technology to other databases. In addition, an actual operational database could be developed, rather than the present semi-realistic simulation.

4.1.9 VALUE OF DIRECTION EXTERNAL

The objective of this experiment is to determine the value of the direction external. This would show the operational value of this parameter, since the current field system does not provide it to a radio station listener.

General Experimental Methodology: Select external set (except location, see Chapter 3). Run several different internal simulations with varying numbers of speakers, languages, and platforms. Run some experiments with ELINT. Evaluate performance for each iteration with and without directional data.

Final Result: The value to the E/I correlation algorithm is X% when directional information is provided along with the other externals.

4.1.10 ELINT/EXTERNAL DATA FUSION

The objective of this experiment is to determine the value of fusing externals with ELINT. The unique ELINT, external, and internal simulation capability provided by this Testbed allows for this comparison to be performed.

General Experimental Methodology: Run several different internal simulations with a varying number of speakers, languages, and platforms. Evaluate performance for each iteration with and without ELINT information.

Final Result: The amount of improvement/degradation due to the merging and correlation of ELINT data with externals and internals is X%.

4.2 EVALUATION

The above ten experiments were analyzed to determine which experiment to run in detail. The factors examined were operational relevance, interests to the speech

community, ease of use (which included setup time), and the experiment's interest to me. The results are summarized in Table 4.

EXPERIMENT NAME	COMPARATIVE OPERATIONAL RELEVANCE	COMPARATIVE INTEREST TO SPEECH COMMUNITY	COMPARATIVE EASE OF USE	COMPARATIVE INTEREST TO ME
1. CORRELATOR 3 PERFORMANCE	MEDIUM	MEDIUM	HIGH	MEDIUM
2. CORRELATOR 3 VS. 4	NOT EVALUATED			
3. BEST EXTERNAL SET	HIGH	LOW	HIGH	MEDIUM
4. OPTIMAL PARAMETER SET	MEDIUM	MEDIUM	HIGH	HIGH
5. LIMITED TRAINING DATA	MEDIUM	MEDIUM	VERY LOW	LOW
6. CORRELATOR 1 VS. 2 VS. 3 VS. 4	NOT EVALUATED			
7. LARGE POPULATION SPEAKER ID	MEDIUM	HIGH	LOW	LOW
8. NEW E/I DATABASE	MEDIUM	MEDIUM	LOW	LOW
9. VALUE OF DIRECTION EXTERNAL	HIGH	LOW	VERY LOW	LOW
10. ELINT/EXTERNAL DATA FUSION	HIGH	LOW	VERY LOW	LOW

TABLE 4: EXPERIMENT TRADE-OFF ANALYSIS

Evaluating correlation algorithm 1, 2, 3, and 4 (Test 6) and comparing Correlator 3 versus 4 (Test 2) were not evaluated. It was preferred to perform more iterations evaluating the performance of Correlator 3 (Test 1), instead of the larger number of iterations and time to effectively prove Test 2 and Test 6.

The experiments with the most operational relevance were the direction (Test 9), ELINT/Externals data fusion (Test 10), best external set (Test 3), and limited training data (Test 5) experiments. Conclusions reached as a result of these tests would influence operational system design.

Tests of data fusion with large population speaker identification (Test 7) is of high interest to the speech processing community. The technical difficulty of separating 300 speakers purely on their speech parameters has been recognized as most researchers

currently only work on groups of 50 or less. It is theorized that somehow breaking the 300 into groups of 50 using externals may be one way of handling the recognition of this large number of speakers.

All the operational tests (Tests 3, 9, and 10) are of low interest to speech researchers except for the limited training experiment. Since the nature of the operational environment is, in general, limited in the amount of audio data for training speech processing systems, researchers would be moderately interested.

Any item was rated very low on the ease of use parameter when is was **required** to run identical data through several different test iterations to prove the conclusion. In Section 5.3.1.2, the impossibility of the Testbed to produce more than one set of reproducible numbers is described.

Any experiment which required modification or creating a new database (Test 7, 8) was rated low on the ease of use criteria. The creation and addition of data files would require an enormous amount of setup time to create the required files. Experiments rated highly on the ease of use parameter are given that rating since the performance of these experiments is trivial once the experiment is designed and the variables are set.

Finally, interest for me was rated as high for the optimization experiment (Test 4). The Testbed was delivered by the contractor with default numerical settings for all parameters. Though some mental thought went into the numerical settings, no optimization was made. Running any of the other experiments with random settings for these parameters would give questionable significance to any results.

Experiment 4 is the most desirable of the options since it scored the best on the total evaluation criteria. In the experimental description in Section 4.1.4, it cited performing experiments to optimize many different parameters. It was later determined that the most important parameters to run experiments with are the merge weights. These weights have a direct bearing on the output scores of the merge algorithm and an indirect bearing on the final hypotheses (since the hypothesis generation algorithm

uses the scores resulting from the merge algorithm).

"Cutoff", "beta", and "gamma" are important parameters only to the hypothesis generation stage of the correlation algorithm. The remaining parameters, other than the merge weights, only have an influence on the Externals and ELINT Processor inputs to the merge algorithm. Though these parameters are important, they are not as important to optimize as the merge weights.

Chapter 5

THESIS EXPERIMENT

This chapter describes in detail the thesis experiment. The chapter contains an introduction, a prediction for the resulting values of the weights, as well as the experimental design, analysis, and conclusions.

In several parts of this chapter, operational values and procedures are cited. This data was obtained from verbal conversations with experienced radio station listeners Norm Lambert of Mei Technology Corporation and Dave Morgan, HRB Systems Inc. The useful operational knowledge of Jim Cupples, Rome Laboratory was also verbally provided.

5.1 INTRODUCTION

The objective of this experiment is to perform tests to optimize the merge weights for ELINT based internals, speech based internals, and external based internals. Since there is no other existing capability to perform this integration of information from these three processes, this experiment will be the initial investigation of the relationships between ELINT, externals, and internals.

One could ask that if one of the processes is consistently better than the others then why run all of them? RL has not done any Research and Development to prove whether internal based speaker identification (ID) is any better than external based speaker ID or ELINT based speaker ID. This holds true for the other internals as well. This is part of the reason for the development of the E/I Testbed.

Internals are only as good as the training. In the operational environment, a multitude of problems (noise, radio station mistuning, speaker variability (e.g., loud, soft,

fast, slow)) cause degradation in speaker ID performance since it is impossible to train the system on all conditions for every speaker.

The E/I contractor declared that external based speaker identification is not as good as internal based speaker identification (Shirey and Morgan, 1993) and thus should be weighted less. However, there are 77 speakers in the database and this large number causes poor external based speaker ID performance. Typical contractor internal simulations used only 10 or 20 speakers, creating a better likelihood of achieving a higher percentage. If there were 77 languages (only 6) and platforms (only 4) in the E/I database the same behavior would be expected for these processes.

While performing experiments, observation showed ELINT based internal recognition percentages in the teens, certainly leading to a theory that ELINT is not a good predictor. This is why in the general experimental methodology in Chapter 4 experiments are recommended to be performed without ELINT.

Despite the fact that each individual process errs, it is necessary to run all of them simultaneously. In internal based speaker ID (as well as language and platform identification) features are extracted then fed into a pattern classification algorithm to make a decision. In several recent RL experiments (Fenstermacher and Smith, 1994) it was noticed that different feature sets through the same classifier erred differently when the decisions were compared. In addition, it was noticed that the same features fed into multiple classifiers also erred differently. Thus, it was theorized that using multiple feature sets and multiple classifiers along with some merging algorithm would provide better results than only using one of them.

This similarity extends to the merging of ELINT, externals and internals as in the E/I Testbed. The three processes all err differently. Thus one expects, as above, that the results provided by the merge would provide more accurate results than each process alone.

5.2 WEIGHT EXPECTATIONS

Based on the overall poor performance of ELINT for each of the internal processes, it is expected to be weighted less than the others. In addition, some experiments are performed without ELINT to compare its performance to experiments with ELINT.

The values for the external weights are not predictable. If the external weights are low, the poor external based speaker identification performance may by de-emphasized at the expense of good external based language identification performance.

It cannot be predicted that the weights for each of the three processes will be equal. RL has seen the internal systems perform differently on different databases. It can only be assumed at this point that the external and ELINT based processes will similarly vary in different parts of the world.

The real answer to this question will be determined when the system is fielded in the future at an operational site, and field data is played into the internals and real externals and ELINT input to the other processes. The key is for representative data to be obtained beforehand to run through the E/I Data Fusion Testbed to insure that the selected weight set will provide "satisfactory" E/I performance (whatever the definition of satisfactory is). This is not necessarily the optimal performance.

This thesis tests whether one weight set is optimal on the semi-realistic simulated data by running 20 iterations of scenario groups (1,2,3) and (4,5,6) and (7,8,9). These groups have different external as well as different internal characteristics. If the "best" weight sets are different, then this gives an indication that there should be different weight sets for different databases.

5.3 EXPERIMENTAL DESIGN

For every experiment, a spreadsheet was designed. This spreadsheet has an experiment name relevant to the experiment and the parameters established for all

variables. An example of this spreadsheet is included in Appendix C. The following sections describe the data on this spreadsheet.

5.3.1 INTERNALS

Real internal outputs could not be input to the Testbed for this experiment. The platform identification algorithms are still in the 6.2 exploratory development stage. Though speaker and language identification software are resident in the RL Speech Processing Facility, the outputs are on a second by second basis not the transmission by transmission decisions required by the E/I Testbed. Some modifications to these algorithms would be required to make their method of reporting compatible with what the Testbed expects.

If all the systems existed and transmission by transmission processing is provided, the Testbed software would be required to handle the outputs from the internals in the interface format provided by the actual systems. This software has not been developed for this interface protocol.

Thus internal simulations were designed, as in the quick experiment in Chapter 3, as realistically as possible. It was desired to run twenty speakers, ten languages, and ten platforms with a percentage accuracy of between 70-75%. These percentages are similar to the actual performance of the RL systems on operational data.

Since the real systems provide confidence scores with *all* their decisions, the number of scores to report was set equal to the number of internals, contrary to the quick experiment in Chapter 3 which only selected five scores from twenty speakers.

5.3.1.1 OBTAINING REALISTIC ID ACCURACIES

There was some design difficulty creating realistic percentage accuracies for these internal simulations. With the speaker ID ground truth score set to 90, only 13% correct ID accuracy was obtained using a random simulation for the secondary scores with no added distortion. When the score was increased to 95, only 41% correct ID accuracy was obtained. After several iterations, a default first choice score of 98 was

finally settled on to give the desired accuracy (68.6%). Similar iterations were performed for language ID and platform ID, establishing a 95 ground truth score.

Though the E/I Testbed provides the capability to use ten languages and platforms, there are only six languages and four platforms in the ground truth database. If language seven (L7) was given a score of 96 by the simulation, the Testbed counted it as an error since it exceeded the 95 language ID ground truth score. However, the merge with zero (since L7 *is not* in the Externals Processor database) resulted in the ground truth language merged score to be higher (since the Externals Processor gave it a score because it *is* in the database). This results in numerous corrections made by the correlation process and a large statistical bias when comparing the merged score results with the simulation results. To correct for this, the simulations for platform and language identification were fixed at the number in the database.

The bias could not be totally removed, however. Due to the simulator definition of "around", whenever L5 and L6 are the ground truth for the language ID simulation, L7 and L8 are generated with scores. This bias is minimized, however, since these two languages occur only 23 times out of 1700 transmissions in the database.

Removing bias for platform ID presented another design difficulty. The platform ID simulation provided an error message when setting the number of platforms to four. The message indicated that the field would only accept numbers between 5 and 20. This was reported to the contractor and it was noted that it was important to change the number of scores chosen to four before changing the number in the database to four.

In addition, it was observed that due to the definition of "around", P5 was simulated whenever P2, P3, and P4 were the ground truth platform. This occurs many times in the database.

After discussions with the contractor, an investigation of constant internal scoring with distortion rather than the above random scoring showed potential for the simulation. Using only language ID, a constant ground truth score of 80 was set with constant secondary scoring of 70 and a random distortion of 11. This produced a

acceptable language ID recognition percentage with scores distributed between 59 and 81.

After extensive analysis, constant internal scoring was preferred as it was determined to be more realistic than purely random scoring. When there is an error made by the RL language identification system, the scores for the two languages being confused are close. In addition, the ground truth score of 80 is more realistic than 95. Finally, all scores are relatively close, simulating a more realistic distribution than when the random simulation gave a spread between 2 and 95.

For each simulation, iterations were performed before an acceptable score spread and simulation accuracy was achieved. This proved difficult, once again, for the speaker identification simulation. The amount of distortion was forced to be as high as possible and the constant secondary score as low as possible. Final values for all simulations are recorded in Appendix C.

5.3.1.2 DYNAMIC VS. REPRODUCIBLE

An investigation into dynamic versus reproducible distortion was performed. It was originally planned to run simulations with and without ELINT, but with the same internals and scores reproduced. It was also originally planned to reproduce the same internals and score sequences through several different iterations of weights.

In this investigation it was discovered that reproducible distortion only produces one sequence of scores over and over again, *not* the same sequence from the last experiment. This was discovered by performing several dynamic and reproducible experiments and observing the language and platform sequences and scores for the first transmission. In every reproducible simulation, only one score sequence was produced. In the dynamic case, all sequences and scores were different.

In the quick experiment in Chapter 3, the statistics for the three scenario groups were different only because of the differing numbers of transmissions between the three scenario groups. The score sequences were exactly the same.

In this experiment the reproducible score sequence will be used for one iteration of the (1,2,3) scenario group. All other experiments will use dynamic scoring. Twenty dynamic iterations of scenario groups (1,2,3), (4,5,6), and (7,8,9) will be run to make the chance of bias less risky. In all cases, the realistic simulation of internals is provided as described.

5.3.2 EXTERNALS

Externals, in the real world, are provided by measurements obtained from the radio station listener using a specially designed system. RL has no in-house access to this system, thus a simulation was used in these experiments. Each scenario group contains different external values, however, all experiments using the same scenario group used the same external values. As described in Section 1.3.2, the simulation of the externals was realistically created based on actual Air Force activities.

Though the preliminary experiment results were not statistically significant, the best external combination of frequency, radio type, and location was chosen as indicated in Appendix C. Frequency was chosen due to its strong ties to the internals and scenarios. Direction and modulation type were not chosen due to the conflicts which appeared to arise in Chapter 3.

Operationally, a radio station listener manually tunes the radio and locks onto a frequency and has direct control of receiving good frequency and location information. In addition the radio type is derived by the knowledge of the radio station listener. Thus, the confidence measures of these externals were kept at the default value of 100.

The Externals Processor relative weights were analyzed and are listed in Appendix C at the system default values (100), since they are only used as multiplicative factors in the algorithm. Changing one value higher or lower would unscientifically favor the contribution of one external over another. The frequency error tolerance was changed to .000lMHz, reflecting a more operationally realistic number.

5.3.3 ELINT

RL has no physical way to recreate the actual ELINT subsystem at present for live input to the Testbed, thus the data is simulated as accurately as possible. The simulation's realism was created by HRB-Systems using a former ELINT analyst.

ELINT Processor relative weights were analyzed and kept at the system default values (100) for the same reason described for the Externals Processor relative weights. The ELINT confidence score was set at 80 showing operationally relevant confidence in the data provided by the ELINT subsystem. The relevant time and distance threshold and error tolerance parameters were kept at operationally significant default values. These values are recorded in Appendix C. In addition, ELINT is designated as "ON" if ELINT is used and "OFF" otherwise.

5.3.4 OTHER EXPERIMENTAL PARAMETERS

Values for "cutoff", "beta", and "gamma" are listed in Appendix C at their default values. Analysis showed that the contractor set these at acceptable values.

The correlation algorithm(3), Externals Processor algorithm(2), ELINT processor algorithm(2) and their respective training files (listed in Appendix C as ASSOCIATIONS.DAT, RELATIONS_ELINT.DAT and RELATIONS.DAT) were kept at the system defaults. The algorithms are claimed by HRB-Systems (Shirey and Morgan, 1993) to be the best. The training files represent 100% of the database, which during testing usually provides the best performance.

Appendix C shows a stop time set at 27000 ms, or 45 minutes which is the entire length of the simulations to provide as much data as possible for analysis. It also shows a run speed of MAX, indicating the experiments will be run as fast as possible. In addition the third flight path (C) was chosen which, as indicated in Chapter 4, does not affect these results since the direction external is not used.

As indicated in Section 5.3.1.2, scenario groups (1,2,3), (4,5,6), and (7,8,9) are being used in these experiments. These groups are indicated in Appendix C after the

CW in the experiment name at the top of the page as well as individually.

Finally, the maximum transmission length was set at 3 seconds. An experiment was performed and concluded that no matter where this parameter was set, the E/I Data Fusion Testbed gave the same results, but provided the information onto the screen faster.

5.3.5 CORRELATION WEIGHTS

For each experiment, the correlation merge weights are listed in the appropriately labeled portion of Appendix C. In every row, the weights are all equal to keep the number of different experiments needed to a reasonable number. This scheme is justified since each internal weighting is independent of each other in the merge algorithm.

Each experiment will change these weights as described in Section 5.3.5.1 and these alterations are recorded on the spreadsheet in Appendix C. The experiment name at the top of the Appendix also includes the weights in the order internals/ externals/ ELINT.

5.3.5.1 EXPERIMENT SERIES

This section describes the series of experiments planned to measure the variability caused by changing the weights of the input processes to the merge algorithm and their effect on the correlation results.

The list of experiments is provided in Appendix D using the form of the experiment name described in Section 5.3.5. The table involves all combinations of weights in intervals of 25. This interval was determined as a tradeoff between the amount of spacing and a reasonable amount of time and number of experiments.

As described in Section 5.1, in this experimental design internals and externals will always be produced (never 0) for correlation. Due to its poor performance, ELINT is not used in some experiments.

It was proven in a short experiment that with ELINT turned off with a non-zero ELINT

weight changes the final results when compared to a zero ELINT weight. This was reported to the contractor as an error, since the software should have automatically made the ELINT weight zero whenever ELINT was off. All experiments without ELINT were performed with a zero weight.

With a short experiment, it was proven that experiments with all weights of 100 provided identical answers when all three weights are set at 75. All experiments whose ratio of weights are multiples of previously listed experiments were not performed and are identified in Appendix D.

Experiments in Appendix D are reproduced for the scenario groups (4,5,6) and (7,8,9) with their experiment names coded appropriately.

5.3.6 ROUTING TABLE

Based on the scenario groups, a routing table was established. The objective of the routing table was to assign the radio stations of interest to the appropriate listener as described in Section 1.1.3.

Operationally, the initial search criteria is language, as this routing criteria sends the appropriate stations to a listener with special education and training in understanding the language. The second most important search criteria is activity, as some news reports are more critical than others. Finally, the last important search criteria is speaker, since known news reporters provide more information than disc jockeys.

Since twelve listeners has operational relevance, an initial attempt at routing was to establish twelve different routing assignments for each of the three scenario groups. It was initially desired that each scenario group have its own routing table.

Initially, each three scenario combination was examined for all possible languages and activities of interest. Attempting to provide for twelve positions with only six language/activity pairs for the scenario (1,2,3) group was unrealistic, so it was decided to only use six listeners. Typically, there are more language/activity pairs and more than

three scenarios.

Only three important language/activity pairs were detected in the scenario (4,5,6) group. These activities could have been divided among two different listeners to cover the desired six positions, but this is operationally unrealistic. Due to this scarceness of the language/activity pairs, it was decided to create *one* realistic twelve position routing table to be used with all three scenario groups.

The final output is contained in Appendix E. The Appendix shows the 12 listeners (P1-P12) with their routing responsibilities coded in the following format: scenario number/language number:speaker number:activity number.

This routing table was prepared by initially creating the language/activity pairs then assigning each pair to a unique listener. No one listener monitors for two different activities in the same scenario group. No one listener monitors for more than one language. In addition, P4 not only listens for A1 in the scenario group (1,2,3) but also in the scenario group (7,8,9).

Finally, representative speakers were added to the language/activity pairs. These speakers represent the most active speakers in the language/activity pairs. The E/I Data Fusion Testbed default routing table was replaced by the configuration specified in Appendix E and used for all experiments.

5.3.7 RECORDING STATISTICS

Appendix F shows the audit file statistics recorded for analysis for each iteration of each experiment. The internal simulation accuracies are recorded as well as the external based activity results. Improvement by the merge algorithm is automatically calculated in the spreadsheet by subtracting these numbers from their respective recorded merge algorithm accuracies. Similarly, correlation improvements are calculated in the same manner using the respective correlator hypotheses, as in the experiment in Chapter 3.

In the correlator hypotheses section, the percent of time the top LSAP hypothesis

was correct was recorded. The percentage of time any of the multiple LSAP hypotheses was correct was recorded and is normalized by the number of hypotheses generated, since all experiments produce a different number of hypotheses.

Some of the statistics used for analysis are labeled with either a direct or indirect relevance. For instance, when optimizing the weights in the merge algorithm, the merged scores are calculated independently of each other and have direct relevance. Since the correlation hypotheses and routing statistics are calculated using the merged scores *and* some internal-internal correlation, it is less likely that these results can be attributed to the quality of the merge weights, thus are indirectly related.

5.4 EXPERIMENTAL PERFORMANCE

Experiments were performed on the Testbed in an automated fashion. The three separate scenario groups and the fourth repeatable group were established with the values described in Section 5.3. All four groups had a version with and without ELINT, creating eight separate experiment files.

In order to change the weights for each experiment listed in Appendix D, the eidefaults file was edited. Editing was also done of the command file to perform experiments and store results for each of the four scenario groups. Initially, a specially designed E/I Editor was used. Later, the Testbed's Openwindows text editor was leaned and used to provide a quicker means to change the weights and create additional experimental data.

The Testbed saved the experimental data and step-by-step results in an audit file for later review and statistical recording. A method to create a short version of this audit file existed on the Testbed. However, after automatically running through several weight groups and scenario groups with this short version, an error was detected in the generated statistics. All iterations processed in this fashion were rerun in the memory intensive step-by-step fashion.

Memory problems on the Testbed resulted after saving several long step-by-step

audit files. As a result, additional experiments had to be performed for those which no memory existed to save the results. Using the UNIX Head and Tail functions in the command file eliminated this problem by only saving the first (to verify the experimental conditions) and last (to print the audit file statistics) 100 lines of each audit file.

To create the data for analysis, these audit file results were recorded. Initially, the results were hand-written onto paper copies of Excel spreadsheets created for each scenario group. This process proved to be slow and tedious. Eventually the statistics from the audit file were printed for each scenario and weight group using the Openwindows Print Monitor.

All printed statistics for each scenario group and weight group were typed into Microsoft Excel spreadsheets illustrated in Appendix F. These spreadsheets were configured to automatically generate the required means and standard deviations for analysis.

Following generation of all spreadsheets, the data was analyzed for accuracy using the standard deviation statistic. After the observation of several spreadsheets, any deviation statistic greater than two was rechecked. This resulted in the correction of several typographical errors resulting from the manual data entry procedure.

Other errors due to manual data entry were detected and corrected in the language identification improvement formulas, the activity correct statistics, and the routing statistics. Since the external and ELINT data was never altered during the 20 iterations of each scenario group, the merge algorithm's activity correct statistic was checked to be constant for all iterations. Due to the way the routing table was created, the mean routing partially correct and the mean routing accuracy statistics were checked to be equal.

5.5 EXPERIMENTAL ANALYSIS

Final statistics generated by the experiments and used for analysis are contained in Appendix G. These statistics illustrate the average improvements and routing

percentages obtained after the merge and correlation algorithms for the 20 iterations of scenario groups (1,2,3), (4,5,6), and (7,8,9). The final statistics also include these same results for the one reproducible iteration of scenario group (1,2,3). The merge and correlation algorithm statistics were summed to create a number signifying the total improvement gained by these two algorithms.

Ranks were created separately for the results of the merge and correlation algorithm using the Microsoft Excel sort function. These ranks were predominantly used in the analysis stage. Though the data in Appendix G was created as carefully as possible, since there was no way to reproduce the same twenty iterations of internal data through all weight groups, the individual statistics are subject to error. However, it is later shown in this thesis that despite these errors, the ranking results for the twenty random iterations of scenario group (1,2,3) are comparable with the one repeatable iteration of (1,2,3) supporting the use of these ranks for this analysis and the conclusions made in this experiment.

In the preliminary analysis stage, the merge algorithm statistics, the correlation hypothesis results, the correlation internal improvements and correlation routing statistics were separately ranked as illustrated in Appendix H. As can be noted by the ranks for the correlation hypothesis results and the routing statistics for weight group 75/50/0, these rankings were sometimes very contradictory within the scenario group and these contradictions made it very difficult to come up with definite trends. Contradictions among scenario groups were also just as plentiful.

To alleviate this analysis a correlation total was calculated and ranked as illustrated in Appendix G. Intuitively, this makes sense since the sum of the correlation hypothesis results, internal improvements, and routing statistics accurately reflects the quality of this algorithm.

For each column in Appendix G, a grand total is calculated for each of the weight groups by summing the merge and correlation totals. This grand total reflects the overall improvement score for a particular weight group.

Using the Microsoft Excel sort function, ranks from best to worst were created for the merge total, correlation total and grand total. It was discovered that when adding the merge and correlation totals in this fashion, the ranking for the grand total was strongly biased in favor of the correlation total since it was generally 7 times larger than the merge total. In order to lessen this bias, the grand totals in Appendix G are calculated by multiplying the merge total by 7, then adding the correlation total.

Finally, the sum of the merge and correlation algorithm ranks was created and separately ranked and this ranking is the way the results are displayed in Appendix G. Once the merge statistics were multiplied by seven, this ranking and the ranking for the grand total are almost identical.

Data analysis was done in three different ways with the results shown in Tables 5, 6, and Figures 3-8. This data is used to perform the comparison of weights in Section 5.5.1.

The first method of analysis began with a judgment of each of the individual numbers in Appendix G. Each row of each scenario group was analyzed separately and significant negative numbers and other deviations noted. An example of a deviation is for scenario group (1,2,3) noting that the statistics obtained when the internal weight is 25 performed worse than all other internal weights.

The correlation algorithm LSP improvement statistic was not analyzed since when the individual LSP statistics were noted as deviant, the total LSP was similarly affected. In addition, the routing statistics completely correct, partially correct and accuracy were not examined since the differences from experiment to experiment were small.

Table 5 gives the results of this first analysis. In this table, the weight groups with nothing noted are not listed. Deviant or negative statistics for scenario group (1,2,3) are denoted as A, repeatable group (1,2,3) are R, scenario group (4,5,6) are B and scenario group (7,8,9) are C.

It was also generally noted while performing this analysis, that as the amount of ELINT decreased, then the results seemed to improve. Another general trend noted at

	100/100/100	100/25/0	100/25/100	100/25/50	100/25/75	100/50/100	100/50/50	100/75/100	25/100/0	25/100/100	25/100/25
MERGE ALGORITHM											
LANGUAGE IMPROVEMENT											
SPEAKER IMPROVEMENT									A,R,B,C	A,R,B,C	A,R,B,C
ACTIVITY IMPROVEMENT					B						
PLATFORM IMPROVEMENT									B	B	B
CORRELATION ALGORITHM											
TOP HYPOTHESIS CORRECT			A,R,B							C	
NORMAL. ANY HYP CORRECT		B	A,R,B	A,R,B			B		B		
LANGUAGE IMPROVEMENT	C		B			C		C			
SPEAKER IMPROVEMENT	C		A,R,B,C	C	B,C	C		C		C	B
ACTIVITY IMPROVEMENT	C		A,R,B,C	C	B,C	C		C			
PLATFORM IMPROVEMENT			A,R,B		C				B		
ROUTING EFFICIENCY		B	A,R,B	A,R,B			B		B		

	25/100/50	25/100/75	25/25/100	25/25/75	25/50/100	25/50/75	25/75/0	25/75/100	25/75/25	25/75/50	25/75/75
MERGE ALGORITHM											
LANGUAGE IMPROVEMENT											
SPEAKER IMPROVEMENT	A,R,B,C	A,R,B,C	A,R,B,C	A,R,B,C	A,R,B,C	A,R,B,C	A,R,B,C	A,R,B,C	A,R,B,C	A,R,B,C	A,R,B,C
ACTIVITY IMPROVEMENT				B							
PLATFORM IMPROVEMENT	B	B			B	B	B	B	B	B	B
CORRELATION ALGORITHM											
TOP HYPOTHESIS CORRECT	C		C	C				C			
NORMAL. ANY HYP CORRECT				A,R							
LANGUAGE IMPROVEMENT	C	A,R	A,R,B,C	B,C	B						
SPEAKER IMPROVEMENT	B,C	A,R,B,C	A,R,B,C	B,C	B	C		A,R,B,C			
ACTIVITY IMPROVEMENT	C	A,R	A,R,B,C	A,R,B,C	B	C		R		C	C
PLATFORM IMPROVEMENT		A,R	A,R,B,C	B	B						
ROUTING EFFICIENCY				A,R							

	50/100/0	50/100/100	50/100/75	50/25/100	50/25/75	50/50/100	50/50/75	50/75/0	75/25/100	75/25/25	75/25/50	75/25/75	75/50/100
MERGE ALGORITHM													
LANGUAGE IMPROVEMENT													
SPEAKER IMPROVEMENT			C	C		C							
ACTIVITY IMPROVEMENT					B							B	
PLATFORM IMPROVEMENT	B	B						B					
CORRELATION ALGORITHM													
TOP HYPOTHESIS CORRECT				C									
NORMAL. ANY HYP CORRECT				A,R,B	A,R,B	R			A		A,R	A,R	
LANGUAGE IMPROVEMENT		C		C			C						
SPEAKER IMPROVEMENT		C	C	B,C	B	C	C		C	C		B,C	C
ACTIVITY IMPROVEMENT		C	C	A,R,B,C	A,R,B	R,C	C		C	C	A,R	A,R,B,C	C
PLATFORM IMPROVEMENT		C		B,C	B							B	
ROUTING EFFICIENCY				A,R,B	A,R,B	R			A		A,R	A,R	

TABLE 5: ANALYSIS OF POOR PERFORMERS

this time was the poorer performance of the correlation language, speaker, activity and platform improvements compared with the results of the merge algorithm.

A second analysis was done by listing the best seven (10% of the 66 total) and worst seven weight groups based on the merge rank and total rank for the four scenario groups. Since the correlation rank is not directly a measure of the alteration of the merge weights (some internal-internal correlation and alternate hypothesis generation occurred), its trends were not examined separately. No analysis of the rank for the final totals in Appendix G was performed since this ranking closely follows the ranking of the sum of the merge and correlation algorithm ranks. Table 6 illustrates the results of this analysis.

The third method of analysis compared accumulated total ranks. For example, to examine the effect of changing the ELINT weight 100/100/100 was placed in a spreadsheet side by side with 100/100/75, 100/100/50, 100/100/25, and 100/100/0. Next, 100/75/100 was placed side by side with 100/75/75, 100/75/50, 100/75/25, and 100/75/0. This continued for all combinations of the internal and external weights.

Following creation of this spreadsheet, the merge ranks and total ranks with the same ELINT weight were added together. Graphs were made for analysis and comparison illustrating the sum of the ranks for each ELINT weight. Graphs were also produced for internals and externals in the same manner and are included in Figures 3-8.

5.5.1 COMPARATIVE DATA ANALYSIS

The similarity of results for scenario group (1,2,3) and the one repeatable iteration of (1,2,3) is verified by the data analysis. In Table 5, A and R are listed in the same column 46 times as opposed to either being listed alone 6 times. In Table 6, the same three weight combinations appear in 23 of 28 columns. Finally, the weight values in Figures 3-8 are extremely close, sometimes exactly overlapping, further supporting this conclusion.

MERGE RANK	BEST (1,2,3)	BEST R(1,2,3)	BEST (4,5,6)	BEST (7,8,9)
1	100/25/0	100/75/0	75/25/50	75/25/0
2	100/100/0	75/50/0	100/75/25	100/25/0
3	75/25/0	100/50/0	100/50/25	75/50/0
4	75/50/0	75/25/0	75/75/25	100/50/25
5	75/50/25	100/25/0	100/50/50	75/75/25
6	100/50/0	100/100/0	75/25/25	100/75/50
7	100/50/25	75/75/25	100/75/50	75/25/25

TOTAL RANK	BEST (1,2,3)	BEST R(1,2,3)	BEST (4,5,6)	BEST (7,8,9)
1	75/50/25	100/75/0	75/75/25	75/50/0
2	75/50/0	75/50/0	75/25/50	75/75/25
3	100/100/25	100/100/25	100/50/25	100/100/0
4	100/100/0	75/75/25	100/75/75	75/25/0
5	75/100/25	75/100/25	50/50/75	100/75/0
6	75/75/25	100/50/0	100/50/100	100/50/25
7	75/100/0	75/50/25	75/50/50	100/50/0

MERGE RANK	WORST (1,2,3)	WORST R(1,2,3)	WORST (4,5,6)	WORST (7,8,9)
60	25/75/50	25/100/75	25/75/50	25/100/100
61	25/100/100	25/75/50	25/75/75	25/100/25
62	25/100/75	25/100/100	25/100/50	25/50/75
63	25/50/75	25/75/100	25/100/100	25/75/50
64	25/25/100	25/50/75	25/25/75	25/50/100
65	25/75/100	25/25/100	25/75/0	25/100/75
66	25/50/100	25/50/100	25/100/0	25/75/100

TOTAL RANK	WORST (1,2,3)	WORST R(1,2,3)	WORST (4,5,6)	WORST (7,8,9)
60	50/25/100	50/25/100	25/100/50	25/25/75
61	25/50/75	25/100/50	25/100/0	25/100/75
62	25/50/100	25/50/75	25/100/25	50/25/100
63	25/25/75	25/25/75	25/50/100	25/100/100
64	25/25/100	25/25/100	25/100/100	25/100/50
65	25/75/100	25/100/75	50/25/75	25/25/100
66	25/100/75	25/75/100	25/25/75	25/75/100

TABLE 6: BEST/WORST ANALYSIS

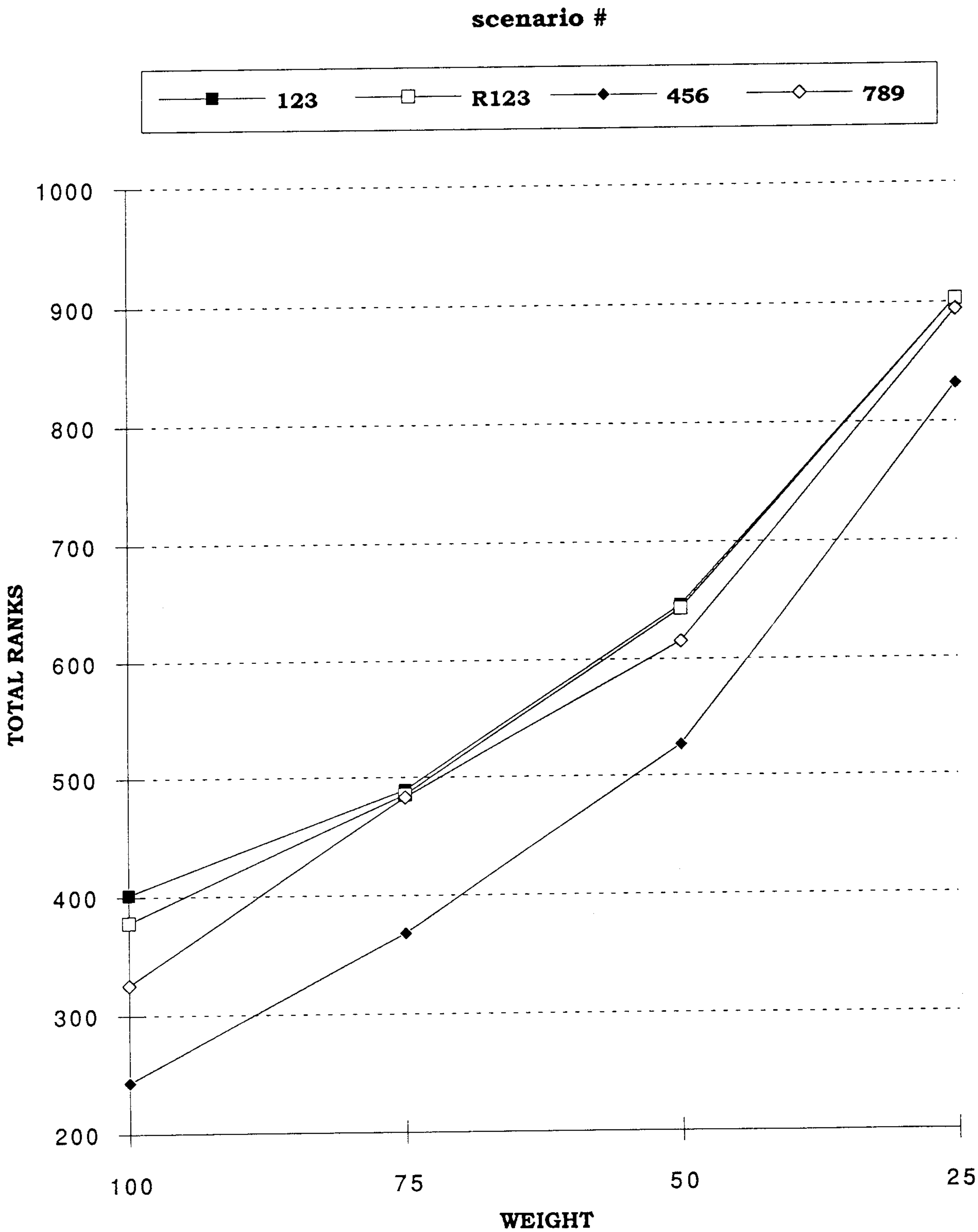

FIGURE 3: GRAPHS OF TOTAL INTERNAL MERGE RANKS

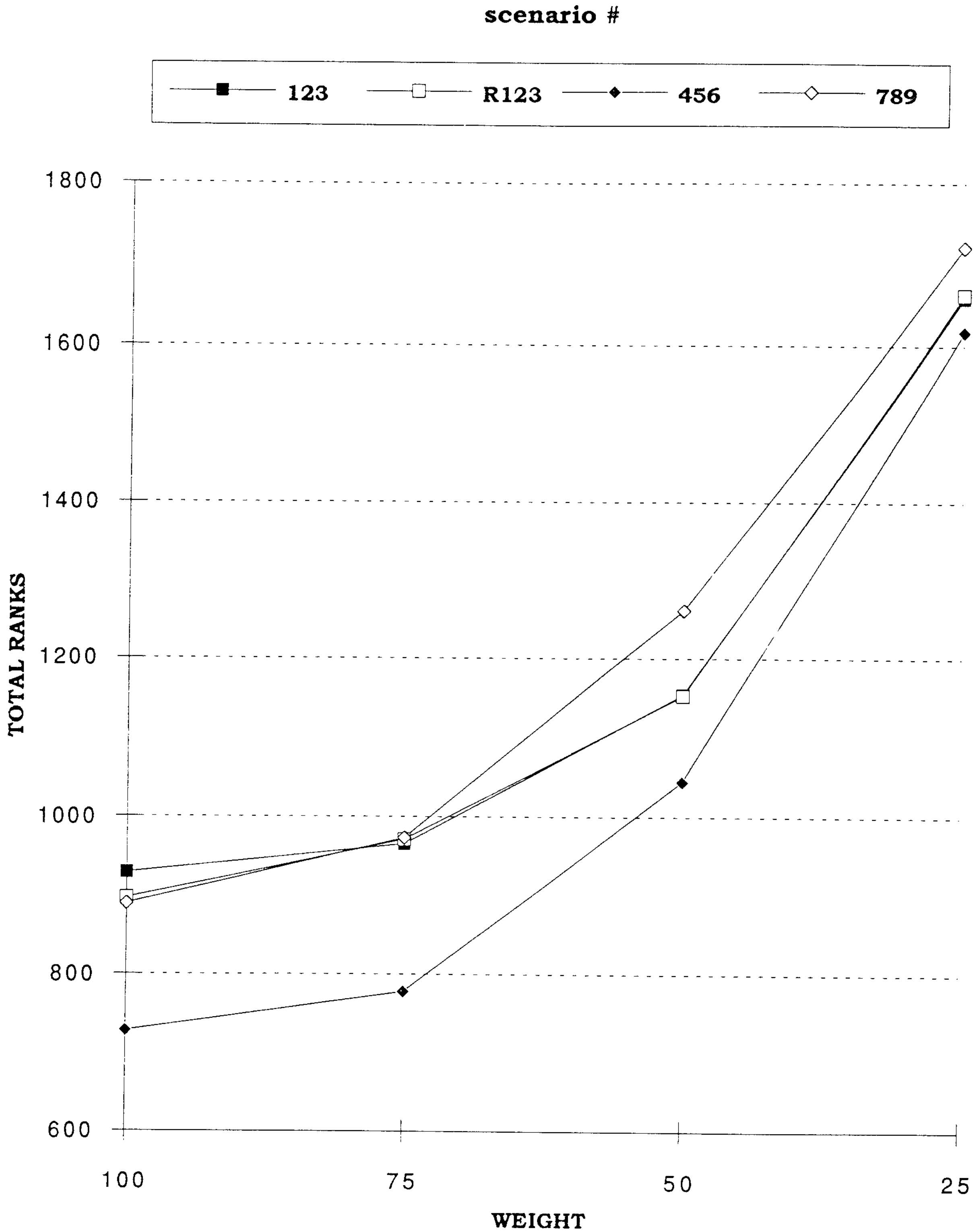

FIGURE 4: GRAPHS OF THE SUM OF THE INTERNAL TOTAL RANKS

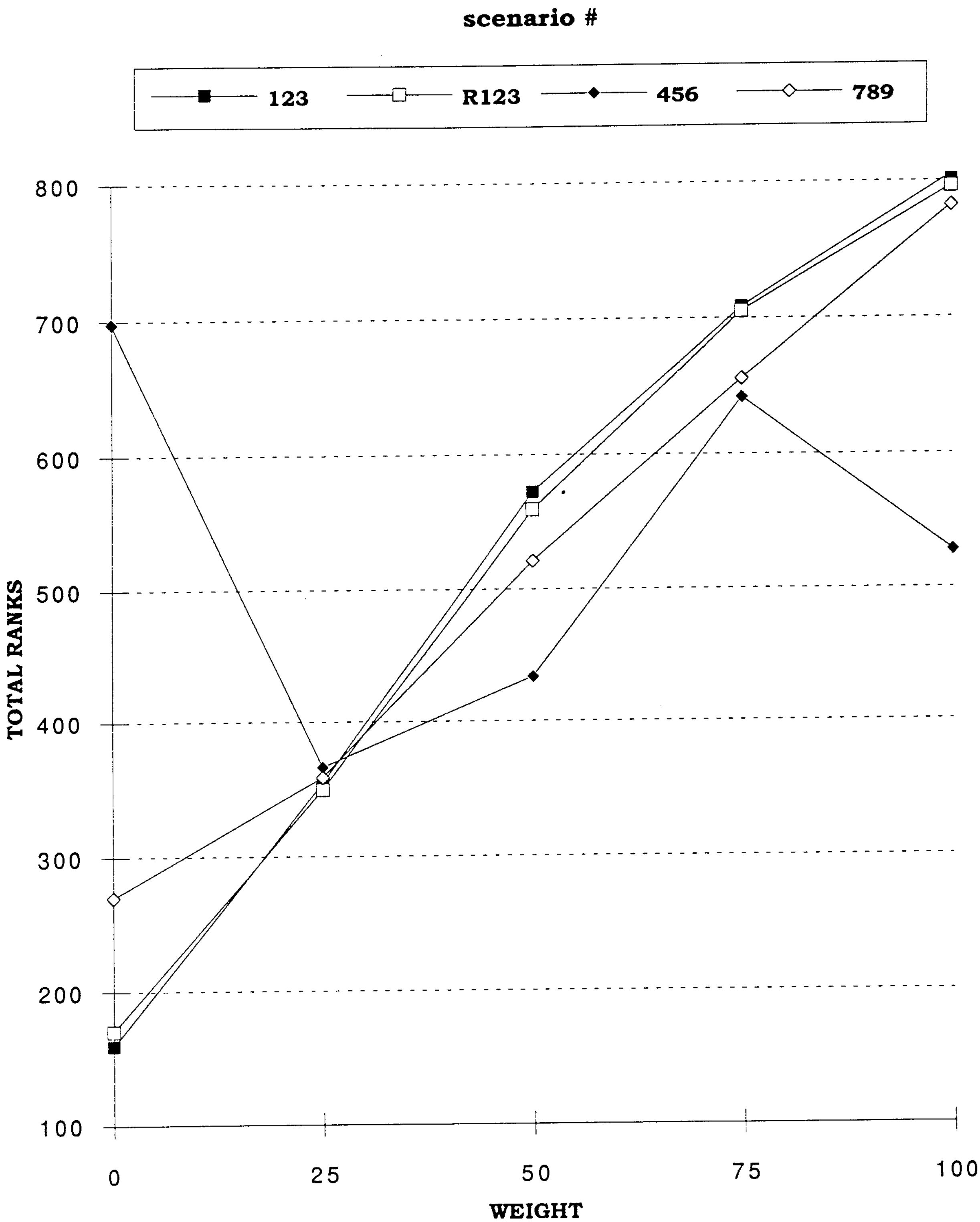

FIGURE 5: GRAPHS OF THE TOTAL ELINT MERGE RANKS

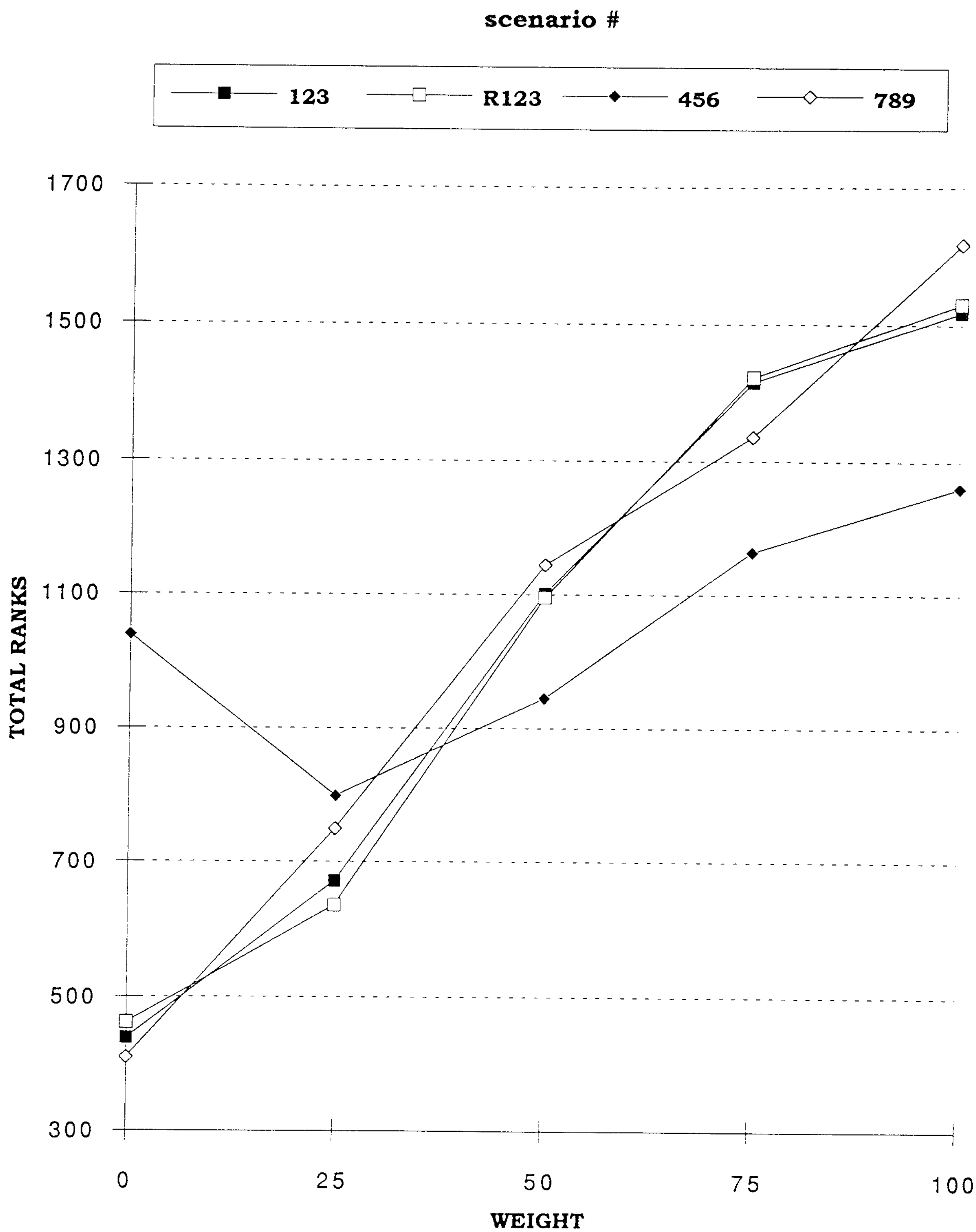

FIGURE 6: GRAPHS OF THE SUM OF THE ELINT TOTAL RANKS

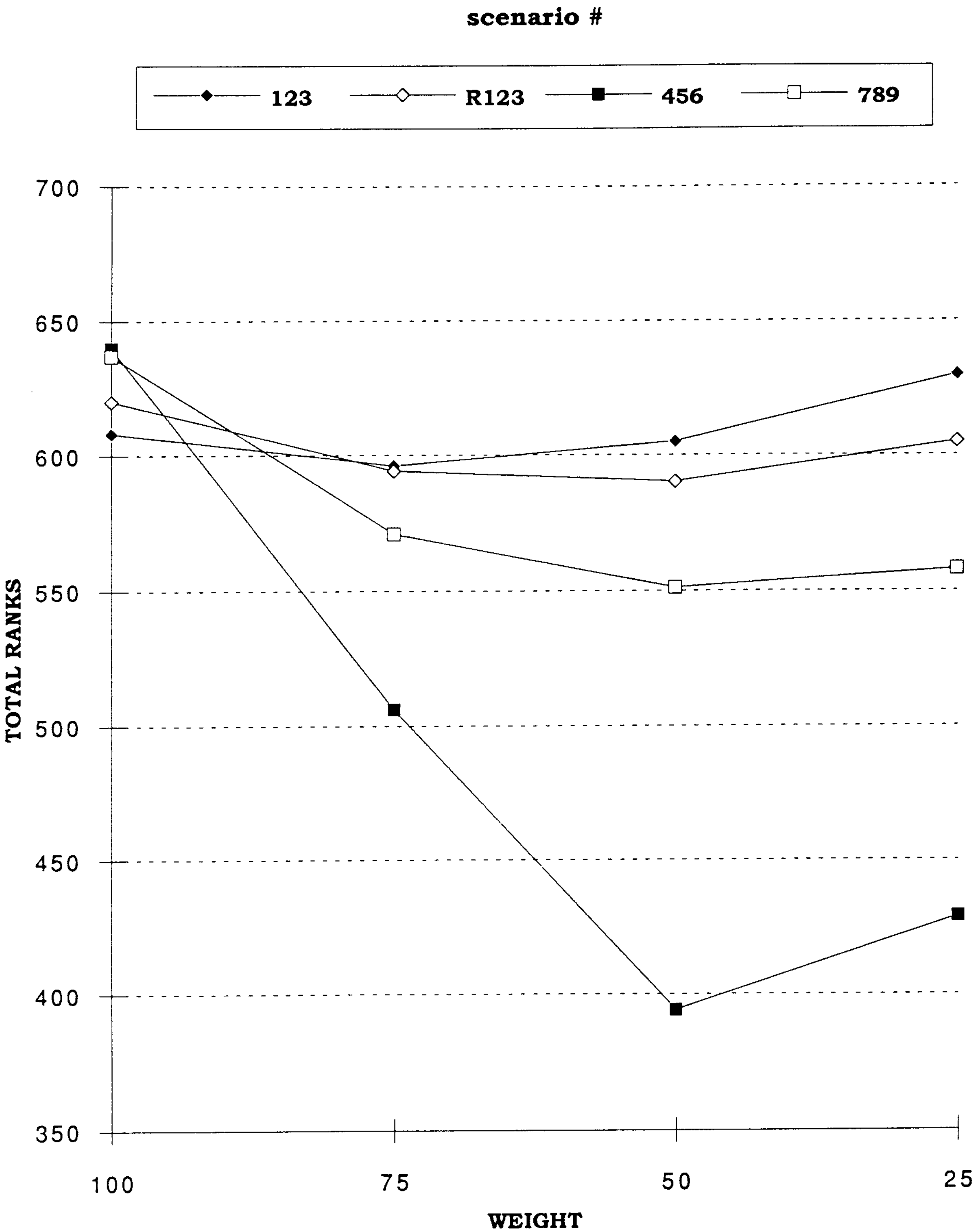

FIGURE 7: GRAPHS OF THE TOTAL EXTERNAL MERGE RANKS

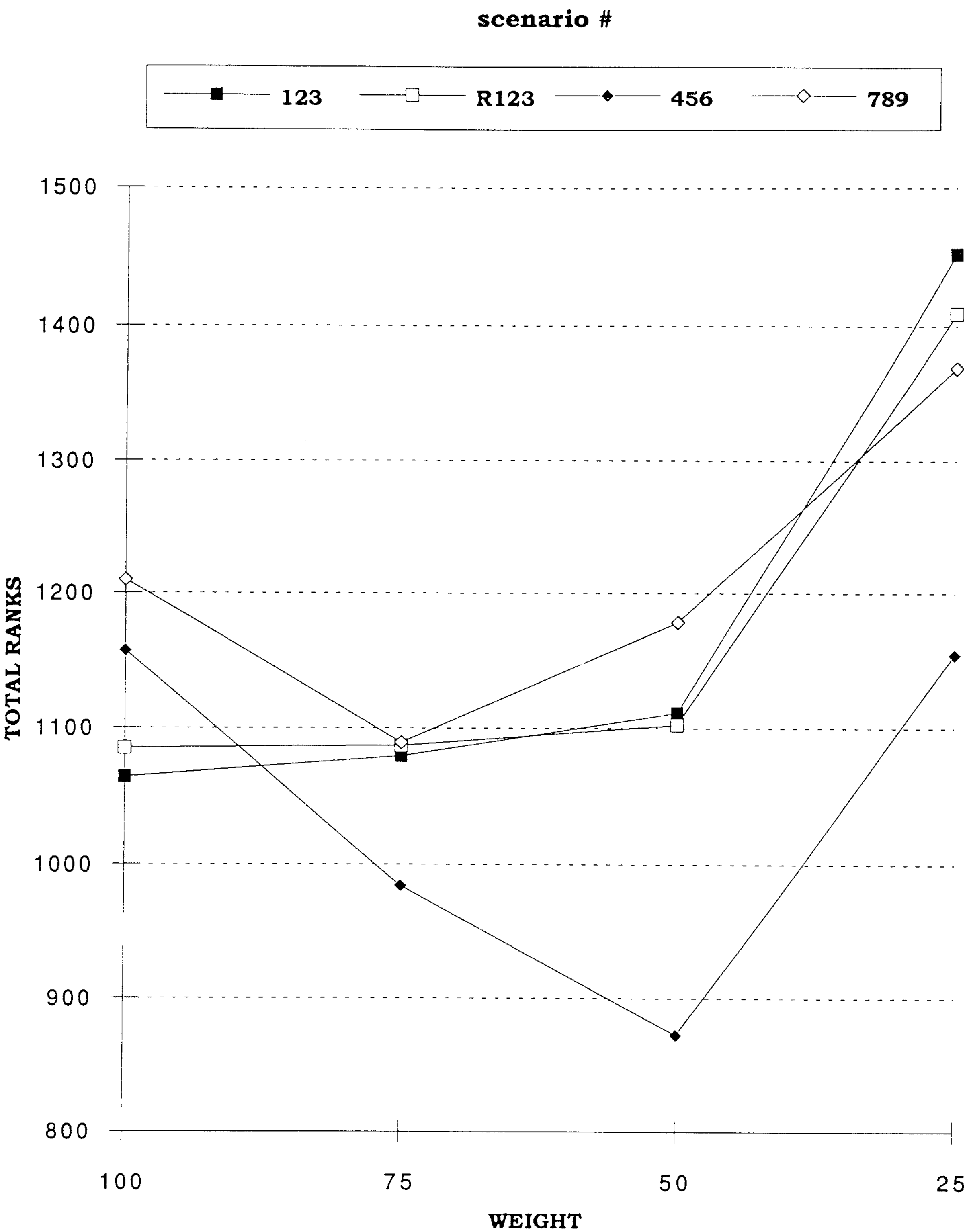

FIGURE 8: GRAPHS OF THE SUM OF THE EXTERNAL TOTAL RANKS

The analyzed data tables and figures were used, where possible, to determine the best and worst weights for internals, externals, and ELINT. Though the overall goal was to find the best, it was important to point out those values which should <u>never</u> be considered to be the best. In some cases, the best or worst analysis was initially narrowed to two or three choices, which were then compared and conclusions made.

It is concluded that having a weight of 25 for the internals is unacceptable for all four scenario groups. In Table 5, this is illustrated by being the internal with the greatest number of letters A, B, C, and R. In addition, all of the combinations in the "WORST" columns in Table 6 except four have an internal weight of 25. Finally, Figures 3 and 4 show much higher total ranks for this weight than any other.

Data analysis also supports the conclusion that an internal weight of 75 or 100 is the best. In Table 5, there are very few occurrences of these internal weights and in most cases the ELINT weight values contribute significantly to these poor results. For example, in the scenario groups labeled A, C, and R, the ELINT weights are high in all cases (50, 75, 100) with one exception (75,25,25). It can be observed in Figures 5 and 6 that these ELINT weights contribute to poorer performance by their higher total ranks for these groups. In addition, for the scenario group labeled B the same thing can be shown with two exceptions (weight combinations 100/50/50 and 100/25/50).

Table 6 also illustrates that best results are achieved with internal weights of 75 or 100. All of the combinations listed as "BEST" in this table contain either of these two weights except for one. This is further reinforced by the low total ranks for these weights in Figures 3 and 4.

There is no strong indication which of these internal weights are better. The numbers appear in the "BEST" columns of Table 6 almost equally (27 versus 28). In addition, Figures 3 and 4 show a lower total rank for the 100 weight, but not by a very significant amount.

The analysis of ELINT weights provides different conclusions for scenario group (4,5,6) as compared with the other scenarios. For group (4,5,6), 25 and 50 are the best,

with 0, 75, and 100 being the worst. For the other 3 scenario groups 0 and 25 are the best, with 75 and 100 being the worst.

For scenario groups (1,2,3), (7,8,9) and the repeatable group (1,2,3), each of the weights judged as worst (75 and 100) individually have many more letters (A,C,R) in the columns of Table 5 than the other weights. In addition, Table 6 has these numbers in the ELINT position in 36 of the 42 "WORST" weight groups. Finally, it can be plainly seen in Figures 5 and 6 that total ranks for these weights are higher than all others for these scenario groups.

As far as which one is the worst value for these scenario groups, it is concluded to be the 100 weight. This weight is in the "WORST" columns of Table 6 twenty-two times compared to fourteen for the 75 weight. In addition, Figures 5 and 6 show the 100 weight consistently with a higher total rank than the 75 weight.

The best weights for these three scenario groups are concluded to be 0 and 25. They are listed only twice in Table 5 without all the other scenario groups and are in every one of the "BEST" ranking combinations in Table 6 except one. Finally, in Figures 5 and 6, ELINT weights of 0 and 25 have the lowest total ranks for these three scenario groups.

The weight of 0 is concluded to be the best for these scenario groups. It occurs more times in Table 6 in the "BEST" columns (25 versus 16) and is the lowest of the total ranks in Figures 5 and 6 making it a better choice than 25.

For the (4,5,6) scenario group, Table 5 shows some activity and platform identification instability in the merge algorithm (illustrated in these rows as a B alone). Further analysis of this table illustrates that there are many B's in the columns with 75 and 100 ELINT weights. In addition, there are a few B's in columns with a zero ELINT weight which are not present for the other groups.

The weight groups in the "WORST" columns of Table 6 for this (4,5,6) scenario group are not dominated by one ELINT weight, though 75 occurs more than any other (4) and there are three occurrences of 100 and 0. Since there are three 50's as well, this table

only reinforces in a small way the poor performance of weights 0, 75 and 100 for this scenario group.

The graphs in Figures 5 and 6 clearly illustrate the poor performance of these weights for this scenario group. These total ranks are much higher than the ranks for the 25 and 5 weights. However, not one of these tables clearly indicates which weight value is worse than the others.

For this scenario group, ELINT weights of 25 and 50 are considered the best. They have the fewest number of B's in Table 5. In Table 6, these values are in all of the "BEST" weight groups in the merge rank table and in four of the seven weight groups in the total rank table, including the top three. Finally, these two weights have the lowest total ranks in Figures 5 and 6.

As for as concluding which of these two weights is better, Table 6 provides no information since they appear an equal number of times in the "BEST" tables. However, Figures 5 and 6 show the weight of 25 with a lower total rank concluding it as the best for this group.

An analysis was performed as to why the zero weight was the best for one scenario group and the worst for the (4,5,6) scenario group. The differences were attributed to the merge algorithm platform and activity instabilities discovered during the analysis of Table 5.

A detailed examination of the merge algorithm platform identification improvement statistic was undertaken for the (4,5,6) group to explain these differences. For this group, the comparative results for the ELINT weight 0 was better than both the 25 and 50 results only 6% of the time. In addition, the results for the 0 weight was better than one of these results only 25% of the time (this result includes the 6%).

This is completely contrary to the results for the other weight groups contributing to the ranking differences in Figures 5 and 6. The repeatable (1,2,3) scenario group always had better results for the 0 weight. For the (7,8,9) scenario group, 50% of the time the results for the 0 weight was better than both the results for the 25 and 50

weights and 75% of the time the 0 weight was better than one of them. Finally, for the (1,2,3) scenario group, 50% of the time the 0 weight was better than both the 25 and 50 weight results and 75% of the time one or the other was better.

The differences in ranking caused by the merge algorithm activity identification statistic were also examined in detail. For the (4,5,6) scenario group, this statistic generally decreased from 5.68 to zero when the ELINT weight is 0. In the other scenario groups, the ELINT weight of zero increased this statistical result by a very small amount. A 5.68 drop in the merge total and final total for the (4,5,6) scenario group equates to about a 15 place increase in rank. Thus, this statistic contributed to the (4,5,6) group's ranking differences.

For externals, a weight of 25 is concluded to be one of the worst values. In Table 5, an A, B, C or R appears more times in these columns than any other external weight. This weight also appears many times in the "WORST" columns of Table 6. In addition, though there is not a significant difference in the ranks in Figure 7, with the exception of scenario group (4,5,6), this weight is never the best. Finally, the ranks in Figure 8 clearly illustrate that this weight is the worst since the total ranks are much higher than any other.

The external weight of 100 is also concluded to be one of the worst values. Though this weight has few letters alone in Table 5, it appears more times than any other external value in the "WORST" columns of Table 6. In addition, it ranked near the poorest in Figure 7. In Figure 8, it appeared as the best for some scenario groups, though not by a significant enough amount to negate this conclusion.

As far as a conclusion for the best external weight, a value of 50 is not listed much in Table 5 and appears the most number of times in the "BEST" ranking columns in Table 6. In addition, in Figures 7 and 8 the scenario group (4,5,6) clearly shows the 50 weight as the best. For the other scenario groups, Figures 7 and 8 show that this weight is either the best or not significantly different from the best.

Considering all three weights at the same time, analysis shows that the weight

combination 75/75/25 is the only combination that appears on all of the "BEST" total rank tables and three of four "BEST" merge ranking tables in Table 6. In addition, further analysis showed that the "BEST" total rank and merge rank weight combinations which provided optimal results in this table (#1 on the list) were never the same.

In most cases, as indicated above, the correlation speaker, language, activity, and platform improvements were less than those provided by the merge algorithm. It is concluded that the convention of generating alternate hypotheses for routing (the star convention as described in Section 2.3) degrades the individual results.

5.6 SUMMARY AND CONCLUSIONS

Tables 7 and 8 summarize the best and worst weights respectively for externals, internals, and ELINT as concluded by these experiments.

	SCENARIO GROUP			
	(1,2,3)	R(1,2,3)	(4,5,6)	(7,8,9)
EXTERNALS	50	50	50	50
INTERNALS	75 OR 100	75 OR 100	75 OR 100	75 OR 100
ELINT	0	0	25	0

TABLE 7: EXPERIMENTAL CONCLUSION: BEST WEIGHT VALUE(S)

	SCENARIO GROUP			
	(1,2,3)	R(1,2,3)	(4,5,6)	(7,8,9)
EXTERNALS	25,100	25,100	25,100	25,100
INTERNALS	25	25	25	25
ELINT	100	100	0,75,100	100

TABLE 8: EXPERIMENTAL CONCLUSION: WORST WEIGHT VALUE(S)

Based on the data for these experiments, there is no unique weight group which is optimal for all four experiment groups. The best scoring combination in each scenario group was not the same and the characteristics of the (4,5,6) group caused its individual weight choices to be different from the others.

Considering the data, the combination 75/75/25 would be a good default for the system. This combination appeared prominently in the "BEST" columns of Table 6. The internal value (75) matches with the value in Table 7. The external value (75) is neither judged best nor worst, however, given the similarities between this weight and the 50 weight in Figures 7 and 8, this value is adequate. Finally the ELINT value was judged as best for scenario group (4,5,6) and a close second best for the other scenario groups making it an adequate choice for the system default.

It is recommended that any future system allow for the evaluation of the effectiveness of routing transmissions based only on the best results of the merge algorithm. Consistently correct LSAP decisions in the nineties were obtained after the merge and good routing statistics are thus expected for the best LSAP combination. If successful, this would alleviate the need for the correlation algorithm which degraded the individual results, but maximizes the routing percentages through the generation of alternate hypotheses.

It is recommended that a capability to reproduce one scenario combination for many experimental conditions be added to the system. Currently, only one reproducible experiment could be run. Adding this capability would allow for more realistic reproducible data to be run through different experimental conditions.

Chapter 6

THESIS SUMMARY

This thesis described the RL Data Fusion Testbed and its development history. All aspects of the Data Fusion Testbed are described in detail including the internals, externals, merge algorithm, and four different correlation algorithms.

This thesis had a heavy emphasis on experimentation. It summarized a group of ten different experiments which could be run using the Testbed and performed a preliminary and detailed experiment.

In the chapter which describes the ten experiments, the objective of each experiment and a general statement of the methodology to run this experiment was given. Finally, expected results are listed.

Each of these experiments was ranked on the following criteria: ease of use, interest to the operational community, interest to the speech processing community, and interest to me. This ranking was used to select the detailed experiment performed.

During performance of this thesis, several computer skills were learned. Among them are several functions from the UNIX operating system, the Openwindows Text Editor and Print Monitor, and a method to list and delete files on the Sun. These skills were used extensively to automatically change weights, run experiments, and record results during the detailed experiment.

This thesis detailed a preliminary experiment which was performed to learn about Testbed experimentation, and to exercise the Testbed simulator and audit file. This preliminary experiment had as its objective to determine which set of externals provides the best correlation results. Experimental procedures during this preliminary experiment were followed during the detailed experiment.

Though the amount of data processed through the preliminary experiment did not result in statistically significant results, the experiment concluded that the combination of frequency, radio type, and location provided the best performance of all external combinations. Performance was measured in terms of mean speaker identification and LSAP (language, speaker, activity, and platform identification) improvement.

The preliminary experiment also showed how different externals and combinations of externals degraded E/I performance. Addition of the direction external to other combinations seemed to hinder performance. The modulation type and radio type externals seemed to work weakly together.

The preliminary experiment also showed a strong correlation between the external frequency and good results since the top scoring combinations included that external. However, it was noted that there were no frequency changes for the scenarios, an operationally unrealistic situation.

The final conclusion of the preliminary experiment examined the performance of the different numbers of externals. This concluded that in accordance with data fusion theory, the larger the number of externals the larger the improvement. The difference between the four and five external case was shown to be smaller than the difference between the other steps.

The detailed experiment had as its objective to analyze the weighting of external, internal and ELINT results. This experiment was carefully designed to reach valid conclusions.

The design of the detailed experiment examined all Testbed parameters in an attempt to be operationally realistic. The design decisions to use twenty iterations, three separate scenario groups ((1,2,3), (4,5,6), and (7,8,9)) and by requiring that the number of platforms and languages be equal to the number available was an attempt to minimize the experimental bias. In addition, a realistic routing table was designed and implemented.

The internal simulations were designed to create operationally realistic accuracies and score spreads. It was concluded through detailed analysis that constant internal scoring with distortion provided more realistic data than random internal scoring.

It was during the design of the internal simulations that it was discovered that there was only one way to reproduce the same data through all different weight groups. Thus, in addition to the three scenario groups, a fourth reproducible group was established.

It was concluded through a very short experiment during the design phase that changing the maximum transmission length parameter only changes the length to time before the testbed displays its results. There is no change in the experimental statistical results.

A list of experiments was created to test the experimental objective. During the design of this list, a short experiment proved that two different experiments where the weights were multiples of each other provided the same results. For example, an experiment with the three processes weighted 50/25/25 produced identical results to an experiment with the weights 100/50/50. This lessened the number of experiments required to be run during the detailed experiment.

Following the performance of all experiments and calculation of the means, Tables 5, 6 and Figures 3-8 were created representing three different ways of looking at the data. During experimental analysis, the results obtained from the one reproducible scenario group very closely followed the performance of the 20 different iterations of the same scenario group for all ELINT, external and internal weights. This conclusion reinforces the conclusions of this detailed experiment, even though there was no way to run multiple iterations of reproducible data through all scenario groups.

Following data analysis the best and worst weights were determined and are repeated in Tables 9 and 10 respectively. These conclusions agreed with the expectations predicted in this thesis.

	SCENARIO GROUP			
	(1,2,3)	R (1,2,3)	(4,5,6)	(7,8,9)
EXTERNALS	50	50	50	50
INTERNALS	75 OR 100	75 OR 100	75 OR 100	75 OR 100
ELINT	0	0	25	0

TABLE 9: BEST WEIGHTS

	SCENARIO GROUP			
	(1,2,3)	R (1,2,3)	(4,5,6)	(7,8,9)
EXTERNALS	25, 100	25, 100	25, 100	25,100
INTERNALS	25	25	25	25
ELINT	100	100	0, 75,100	100

TABLE 10: WORST WEIGHTS

The expectation that the conclusions would be different for each weight group was confirmed. The conclusions for the (4,5,6) group were different than the others due to platform and activity improvement instability.

A 75/75/25 weight group was recommended for the system default for all four scenario groups. Though these weights were not rated as the "BEST" for all scenario groups, analysis showed them adequate enough to be recommended as the default weights.

Analysis also showed that the merge algorithms LSAP improvements were consistently greater than the correlation algorithm improvements. This questions the use of the correlation algorithm.

During performance of these experiments three software errors were detected. When ELINT is not used, the weight for ELINT is not made 0 causing for some statistical differences as compared to when the weight is 0. In the E/I Editor, the search and

replace function would not work. Finally, when trying to run automated experiments in the short fashion, the routing statistics turned out incorrect.

Besides correcting these errors, various recommendations are made through this thesis. Experiments with a database with frequency changes would test actual strength of the frequency external, and would be more operationally realistic then the E/I database.

Changing the software to be able to run experiments with the same data is recommended. This would allow for more controlled experiments, with repeatable results.

Other data fusion and correlation algorithms should be implemented, tested and evaluated against the current ones. In addition, the decrease in performance due to the current correlation algorithm should be analyzed.

Allowing the evaluation of routing based on the results of the merge algorithm is recommended. The percentages for each internal after the merge algorithm was consistently ninety and above which would expect routing accuracies greater than the 40% currently obtained in the field.

Finally, the addition of statistical analysis routines to the Testbed will eliminate the need to do these functions manually as was done in this thesis. Addition of these routines would make experimental data collection faster.

BIBLIOGRAPHY

Anthony, Robert N. et al, *Management Control Systems*, 7th Edition, Homewood, IL: Richard D. Irwin Inc., 1992.

Boone, Louis E. and David L. Kurtz, *Contemporary Marketing*, 6th Edition, Cincinnati, OH: Dryden Press, 1989.

Fenstermacher, Laurie and Douglas Smith, *Tactical Speaker Recognition Using Feature and Classifier Fusion*, SPIE- The International Society for Optical Engineering Proceedings Vol. 2243, April 1994.

* George, James, *External/Internal Processor*, Rome Laboratory Final Technical Report, 1990.

Gitman, Lawrence J., *Principles of Managerial Finance*, 6th Edition, New York, NY: Harper Collins Publishers Inc., 1991.

* Hamer, Wayne and Jerry Foil, *External/Internal Signal Analysis Study*, Rome Laboratory Final Technical Report, 1987.

* Lambert, Norm et al, *Gisting*, Draft Rome Laboratory Final Technical Report, 1993.

Meredith, Jack R. and Samuel Mantel, Jr., *Project Management: A Managerial Approach*, 2nd Edition, New York, NY: John Wiley and Sons, Inc., 1989.

Moore, David S. and George P. McCabe, *Introduction to the Practice of Statistics*, New York, NY: W. H. Freeman and Co., 1989.

Parker, John, *External/Internal Data Fusion: History, Components, Uses, and Proposed Experimentation*, Draft Rome Laboratory Technical Memo, 1993.

Perz, Lt. Mike and John Parker, *External/Internal Data Fusion*, Draft Rome Laboratory Technical Memo, 1993.

Robbins, Stephen, *Organizational Behavior, Concepts, Controversies, and Applications*, 5th Edition, Englewood Cliffs, NJ: Prentice Hall Inc., 1991.

* Shirey, Mark and Dave Morgan, *External/Internal Data Fusion*, Rome Laboratory Final Technical Report, 1993.

Silver, Gerald A. and Myrna L. Silver, *Data Communications for Business*, Boston, MA: Boyd & Fraser Publishing Co., 1987.

Stamper, David and Wilson Price, *Database Design and Management: An Applied Approach*, New York, NY: McGraw-Hill, 1990.

Waltz, Edward and James Llinas, *Multisensor Data Fusion*, Boston, MA: Artech House, 1990.

Weygandt, Jerry J. et al, *Accounting Principles*, 2nd Edition, New York, NY: John Wiley and Sons, Inc., 1990.

* Woodsum, Harvey and Walter Harms, *External/Internal Signal Processor EDM*, Rome Laboratory Final Technical Report, 1987.

*NOTE: Although this report references * limited documents above, no limited information has been extracted.

APPENDIXES

APPENDIX A

SYSTEM DEFAULTS IN EIDEFAULTS FILE

DEFAULT ALGORITHM NUMBER / **TRAINING FILES**

CORRELATION ALGORITHM=	3	ASSOCIATIONS.DAT
ELINT PROCESSOR ALGORITHM=	2	RELATIONS_ELINT.DAT
EXTERNALS PROCESSOR ALGORITHM=	2	RELATIONS.DAT

EXTERNAL SIMULATION

FREQUENCY:	ERROR TOLERANCE=	0.001	RELATIVE WEIGHT=	100
RADIO TYPE			RELATIVE WEIGHT=	100
LOCATION			RELATIVE WEIGHT=	100
DIRECTION			RELATIVE WEIGHT=	100
MODULATION TYPE			RELATIVE WEIGHT=	100

ELINT SIMULATION

ELINT RELEVANT TIME THRESHOLD=	1200	TYPE RELATIVE WEIGHT=	100
ELINT RELEVANT DISTANCE THRESHOLD=	1	LOCATION RELATIVE WEIGHT=	100
ELINT LOCATION ERROR TOLERANCE=	3		

CORRELATION WEIGHTS

	LANG	SPK	ACT	PLAT
INTERNAL SIMULATION	100	100	---	100
EXTERNALS PROCESSOR	100	100	100	100
ELINT PROCESSOR	100	100	100	100

OTHER PARAMETERS

BETA=	3	CUTOFF=	5
GAMMA=	2		

APPENDIX B

SAMPLE OF QUICK EXPERIMENT RESULTS

	5 EXTERNALS				
SPEAKER ID:	(1,2,3)	(4,5,6)	(7,8,9)		
Correlator hypo: spkr. corr.	84.86%	88.33%	78.40%		STANDARD
Simulator: speaker correct	64.94%	66.11%	66.67%	MEAN	DEVIATION
% speaker improvement	19.92	22.22	11.73	17.96	5.51
LSP COMBINATION:					
LSP correlated correct	83.27%	88.33%	78.38%		STANDARD
LSP internals correct	64.94%	66.11%	66.67%	MEAN	DEVIATION
% LSP improvement	18.33	22.22	11.71	17.42	5.31
AVG # hypotheses	4.03	3.83	3.78		
CORRELATOR ROUTING:					
completely correct	94.42%	100.00%	90.12%		
partially correct	100.00%	100.00%	100.00%		
accuracy	97.43%	100.00%	95.52%		
efficiency	92.50%	94.63%	88.34%		

	4 EXTERNALS DIR, MOD, RT, LOC				
SPEAKER ID:	(1,2,3)	(4,5,6)	(7,8,9)		
Correlator hypo: spkr. corr.	80.08%	85.00%	70.37%		STANDARD
Simulator: speaker correct	64.94%	66.11%	66.67%	MEAN	DEVIATION
% speaker improvement	15.14%	18.89%	3.70%	12.58	7.91
LSP COMBINATION:					
LSP correlated correct	78.49%	76.67%	69.75%		STANDARD
LSP internals correct	64.94%	66.11%	66.67%	MEAN	DEVIATION
% LSP improvement	13.55%	10.56%	3.08%	9.06	5.39
AVG # hypotheses	4.77	6.10	4.79		
CORRELATOR ROUTING:					
completely correct	94.42%	100.00%	90.74%		
partially correct	100.00%	100.00%	100.00%		
accuracy	97.43%	100.00%	95.80%		
efficiency	88.78%	83.71%	85.93%		

	4 EXTERNALS FREQ, MOD, RT, LOC				
SPEAKER ID:	(1,2,3)	(4,5,6)	(7,8,9)		
Correlator hypo: spkr. corr.	86.06%	93.33%	82.72%		STANDARD
Simulator: speaker correct	64.94%	66.11%	66.67%	MEAN	DEVIATION
% speaker improvement	21.12	27.22	16.05	21.46	5.59
LSP COMBINATION:					
LSP correlated correct	85.26%	93.33%	82.72%		STANDARD
LSP internals correct	64.94%	66.11%	66.67%	MEAN	DEVIATION
% LSP improvement	20.32	27.22	16.05	21.20	5.64
AVG # hypotheses	2.92	2.56	3.05		
CORRELATOR ROUTING:					
completely correct	94.42%	98.89%	85.80%		
partially correct	100.00%	100.00%	100.00%		
accuracy	97.43%	99.46%	93.56%		
efficiency	98.70%	96.84%	97.95%		

APPENDIX C

EXPERIMENT WORKSHEET

EXPERIMENT NAME:CW123 /25/75/100

CORRELATION ALGORITHM=	3
ELINT PROCESSOR ALGORITHM=	2
EXTERNALS PROCESSOR ALGOR.=	2
STOP TIME=	27000
RUN SPEED =	MAX
FLIGHT PATH =	C
MAX TRANSMISSION LENGTH=	3

TRAINING FILES

ASSOCIATIONS.DAT
RELATIONS_ELINT.DAT
RELATIONS.DAT

INTERNAL SIMULATION

# LANGUAGES & SCORES=	6	# SPEAKERS & SCORES=	20	# PLATFORMS & SCORES=	4
CONSTANT OR RANDOM	C	CONSTANT OR RANDOM	C	CONSTANT OR RANDOM	C
GROUND TRUTH SCORE=	80	GROUND TRUTH SCORE=	80	GROUND TRUTH SCORE=	80
SECONDARY SCORES	60	SECONDARY SCORES	45	SECONDARY SCORES	60
DISTORTION=	23	DISTORTION=	37	DISTORTION=	25

EXTERNAL SIMULATION

FREQUENCY CONFIDENCE SCORE=	100	RELATIVE WEIGHT=	100	ERROR TOLERANCE=	0.0001
RADIO TYPE CONFIDENCE SCORE=	100	RELATIVE WEIGHT=	100		
LOCATION CONFIDENCE SCORE=	100	RELATIVE WEIGHT=	100		

ELINT SIMULATION

ELINT ON/OFF	ON				
ELINT CONFIDENCE SCORE=	80	TYPE RELATIVE WEIGHT=	100		
ELINT RELEVANT TIME THRESHOLD=	1200	LOCATION RELATIVE WEIGHT=	100	ERROR TOLERANCE=	3
ELINT·RELEVANT DISTANCE THRESHOLD=	1				

CORRELATION WEIGHTS

	LANG	SPK	ACT	PLAT
INTERNAL SIMULATION	100	100	---	100
EXTERNALS PROCESSOR	100	100	100	100
ELINT PROCESSOR	100	100	100	100

OTHER PARAMETERS

SCENARIO #	1	SCENARIO#	2	SCENARIO#	3
BETA=	3	CUTOFF=	5	GAMMA=	2

APPENDIX D

EXPERIMENT LIST

EXPERIMENTS WITH ELINT

CW123 /100/100/100
CW123 /100/100/75
CW123 /100/100/50
CW123 /100/100/25
CW123 /100/75/100
CW123 /100/75/75
CW123 /100/75/50
CW123 /100/75/25
CW123 /100/50/100
CW123 /100/50/75
CW123 /100/50/50
CW123 /100/50/25
CW123 /100/25/100
CW123 /100/25/75
CW123 /100/25/50
CW123 /100/25/25
CW123 /75/100/100
CW123 /75/75/100
CW123 /75/50/100
CW123 /75/25/100
CW123 /50/100/100
CW123 /50/75/100
CW123 /50/50/100
CW123 /50/25/100
CW123 /25/100/100
CW123 /25/75/100
CW123 /25/50/100
CW123 /25/25/100
CW123 /75/100/75
CW123 /75/100/50
CW123 /75/100/25
CW123 /50/100/75
CW123 /50/100/50
CW123 /50/100/25
CW123 /25/100/75
CW123 /25/100/50
CW123 /25/100/25
CW123 /75/75/25
CW123 /75/75/50
CW123 /75/50/75
CW123 /75/50/50
CW123 /75/50/25
CW123 /75/25/75
CW123 /75/25/50
CW123 /75/25/25
CW123 /50/75/75
CW123 /50/75/50
CW123 /50/75/25
CW123 /50/50/75
CW123 /50/25/75
CW123 /25/25/75
CW123 /25/50/75
CW123 /25/75/25
CW123 /25/75/50
CW123 /25/75/75

EXPERIMENTS WITHOUT ELINT

CW123 /100/100/0
CW123 /100/75/0
CW123 /100/50/0
CW123 /100/25/0
CW123 /75/100/0
CW123 /50/100/0
CW123 /25/100/0
CW123 /75/50/0
CW123 /75/25/0
CW123 /50/75/0
CW123 /25/75/0

EXPERIMENTS NOT PERFORMED

CW123 /75/75/75
CW123 /50/50/50
CW123 /50/50/25
CW123 /50/25/25
CW123 /50/25/50
CW123 /25/50/25
CW123 /25/50/50
CW123 /25/25/25
CW123 /25/25/50
CW123 /50/25/0
CW123 /25/50/0
CW123 /75/75/0
CW123 /50/50/0
CW123 /25/25/0

APPENDIX E

ROUTING TABLE

P1

4 / L1:I2: A5
4 / L1:B2:A5
4 / L1:C2:A5
4 / L1:A2:A5

P2

2 / L1:H1:A10
2 / L1:J1:A10
2 / L1:I1: A10

P3

5 / L1:G2:A7
5 / L1:F2:A7
5 / L1:O2:A7
5 / L1:L2:A7

P4

1 / L1:A1:A1
1 / L1:P1:A1
1 / L1:N1:A1
1 / L1:M1:A1
8 / L1:A3:A1
8 / L1:B3:A1

P5

7 / L1:H3:A11
7 / L1:G3:A11
7 / L1:C3:A11
7 / L1:M3:A11
7 / L1:L3:A11
7 / L1:K3:A11
7 / L1:F3:A11

P6

3 / L1:G1:A3
3 / L1:B1:A3
3 / L1:C1:A3
8 / L1:B3:A6
8 / L1:P3:A6
8 / L1:O3:A6

P7

1 / L1:A1:A4
1 / L1:B1:A4

P8

2 / L1: I1:A9
4 / L1:D2:A8

P9

9 / L5:V3 :A1
9 / L5:W3:A1

P10

9 / L4:S3:A3
9 / L4:T3:A3

P11

6 / L3:H2:A1
6 / L3:U2:A1

P12

1 / L6:Y1:A1
1 / L6:X1:A1

APPENDIX F

RESULTS TO BE ANALYZED

EXPERIMENT NAME:CW123 /25/75/100 3

INTERNAL SIMULATION

LANGUAGE CORRECT
SPEAKER CORRECT
PLATFORM CORRECT
LSP CORRECT

EXTERNAL BASED

ACTIVITY CORRECT

MERGE ALGORITHM

DIRECT RELEVANCE

LANGUAGE CORRECT	IMPROVEMENT DUE TO MERGING
SPEAKER CORRECT	IMPROVEMENT DUE TO MERGING
ACTIVITY CORRECT	IMPROVEMENT DUE TO MERGING
PLATFORM CORRECT	IMPROVEMENT DUE TO MERGING

CORRELATOR HYPOTHESES

INDIRECT RELEVANCE

	TOP HYPOTHESIS CORRECT
ANY HYPOTHESIS CORRECT	
AVERAGE # HYPOTHESES	NORMALIZED ANY HYPOT. CORRECT
LANGUAGE CORRECT	IMPROVEMENT DUE TO CORRELATION
SPEAKER CORRECT	IMPROVEMENT DUE TO CORRELATION
ACTIVITY CORRECT	IMPROVEMENT DUE TO CORRELATION
PLATFORM CORRECT	IMPROVEMENT DUE TO CORRELATION
LSP CORRECT	IMPROVEMENT DUE TO CORRELATION

CORRELATOR ROUTING

INDIRECT RELEVANCE

COMPLETELY CORRECT
PARTIALLY CORRECT
ACCURACY
EFFICIENCY

APPENDIX G

FINAL RESULTS

SCENARIO GROUP (1,2,3)	75/50/25	75/50/0	100/100/25	100/100/0	75/100/25	75/75/25	75/100/0	75/100/50	100/50/25
MERGE ALGORITHM									
LANGUAGE IMPROVEMENT	28.76	28.65	27.76	29.14	27.80	28.96	28.61	27.37	29.41
SPEAKER IMPROVEMENT	36.05	36.77	36.35	36.61	35.59	36.20	36.48	35.34	36.69
ACTIVITY IMPROVEMENT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PLATFORM IMPROVEMENT	27.03	26.81	26.60	26.75	26.88	26.06	26.27	27.72	25.65
TOTAL	**91.83**	**92.23**	**90.71**	**92.49**	**90.26**	**91.22**	**91.36**	**90.42**	**91.74**
RANK	**5.00**	**4.00**	**12.00**	**2.00**	**15.00**	**9.00**	**8.00**	**14.00**	**7.00**
CORRELATION ALGORITHM									
TOP HYPOTHESIS CORRECT	87.67	88.48	88.08	86.45	90.39	86.58	86.26	86.73	86.90
NORMAL. ANY HYP CORRECT	44.73	41.29	50.92	42.93	55.59	49.46	49.39	50.35	42.28
LANGUAGE IMPROVEMENT	26.88	25.77	25.04	26.14	25.97	25.69	24.96	25.19	27.24
SPEAKER IMPROVEMENT	31.69	31.51	32.09	30.03	33.46	29.85	29.88	31.08	30.32
ACTIVITY IMPROVEMENT	-2.54	-2.16	-2.01	-2.78	-1.68	-2.43	-2.83	-1.84	-2.51
PLATFORM IMPROVEMENT	24.12	23.41	23.27	22.85	24.24	22.12	22.47	24.38	22.77
LSP IMPROVEMENT	60.37	59.98	59.74	58.37	61.67	57.95	57.52	58.90	58.45
ROUTING									
COMPLETELY CORRECT	97.43	97.55	97.44	97.56	97.43	97.55	97.55	97.55	97.55
PARTIALLY CORRECT	97.04	97.19	97.05	97.20	97.04	97.19	97.19	97.19	97.19
ACCURACY	97.04	97.19	97.05	97.20	97.04	97.19	97.19	97.19	97.19
EFFICIENCY	88.42	89.80	89.31	89.52	89.94	89.32	89.59	88.41	88.53
TOTAL	**652.86**	**650.01**	**657.98**	**645.48**	**671.07**	**650.48**	**649.16**	**655.14**	**645.91**
RANK	**5.00**	**9.00**	**2.00**	**13.00**	**1.00**	**7.00**	**10.00**	**4.00**	**12.00**
GRAND TOTAL	**1295.67**	**1295.62**	**1292.93**	**1292.93**	**1302.88**	**1288.99**	**1288.67**	**1288.08**	**1288.11**
RANK FOR GRAND TOTAL	**2.00**	**3.00**	**4.00**	**4.00**	**1.00**	**6.00**	**7.00**	**9.00**	**8.00**
TOTAL RANK	10.00	13.00	14.00	15.00	16.00	16.00	18.00	18.00	19.00

SCENARIO GROUP (4,5,6)	75/75/25	75/25/50	100/50/25	100/75/75	50/50/75	100/50/100	75/50/50	100/50/75	75/50/75
MERGE ALGORITHM									
LANGUAGE IMPROVEMENT	28.73	28.50	28.32	26.29	26.80	27.04	27.09	26.98	27.58
SPEAKER IMPROVEMENT	35.96	36.19	36.09	34.09	35.22	34.73	35.37	34.10	33.89
ACTIVITY IMPROVEMENT	5.68	5.68	5.68	5.68	5.68	5.68	5.68	5.68	5.68
PLATFORM IMPROVEMENT	20.90	21.79	21.65	21.21	19.46	22.90	19.88	23.68	19.46
TOTAL	**91.27**	**92.17**	**91.74**	**87.27**	**87.16**	**90.34**	**88.03**	**90.44**	**86.62**
RANK	**4.00**	**1.00**	**3.00**	**18.00**	**19.00**	**9.00**	**16.00**	**8.00**	**21.00**
CORRELATION ALGORITHM									
TOP HYPOTHESIS CORRECT	87.36	85.79	86.33	89.88	86.12	83.69	85.48	79.82	84.30
NORMAL. ANY HYP CORRECT	36.35	35.00	35.25	48.43	43.90	35.66	39.20	36.90	40.09
LANGUAGE IMPROVEMENT	26.06	25.89	25.18	23.80	24.48	23.23	24.84	24.03	25.53
SPEAKER IMPROVEMENT	29.88	29.08	27.79	33.41	30.14	26.46	28.26	21.78	27.12
ACTIVITY IMPROVEMENT	12.47	9.63	12.16	14.13	12.63	9.00	10.70	3.61	9.08
PLATFORM IMPROVEMENT	24.36	21.78	23.73	24.94	23.21	22.48	22.42	20.38	19.90
LSP IMPROVEMENT	59.34	56.19	56.93	60.55	57.02	54.48	55.52	49.44	53.69
ROUTING									
COMPLETELY CORRECT	99.58	99.26	99.34	98.63	98.87	99.19	99.19	99.20	99.10
PARTIALLY CORRECT	99.53	99.16	99.26	98.45	98.72	99.09	99.09	99.10	98.99
ACCURACY	99.53	99.16	99.26	98.45	98.72	99.09	99.09	99.10	98.99
EFFICIENCY	53.40	55.11	48.00	80.95	70.25	50.20	46.53	54.67	65.09
TOTAL	**627.85**	**616.05**	**613.23**	**671.62**	**644.04**	**602.55**	**610.31**	**588.02**	**621.87**
RANK	**7.00**	**10.00**	**11.00**	**1.00**	**2.00**	**16.00**	**13.00**	**21.00**	**9.00**
GRAND TOTAL	**1266.71**	**1261.21**	**1255.39**	**1282.55**	**1254.19**	**1234.96**	**1226.50**	**1221.10**	**1228.19**
RANK FOR GRAND TOTAL	**2.00**	**3.00**	**4.00**	**1.00**	**5.00**	**6.00**	**9.00**	**10.00**	**7.00**
TOTAL RANK	11.00	11.00	14.00	19.00	21.00	25.00	29.00	29.00	30.00

SCENARIO GROUP (1,2,3)	100/50/0	100/75/0	50/75/25	75/25/0	50/75/0	100/100/50	100/75/25	50/100/25	100/25/0	50/100/0
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	28.07	27.39	28.14	28.98	27.69	28.44	27.94	26.46	29.21	26.08
SPEAKER IMPROVEMENT	36.40	37.91	36.23	38.10	35.60	35.86	36.84	35.69	37.72	36.15
ACTIVITY IMPROVEMENT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PLATFORM IMPROVEMENT	27.33	25.47	24.94	25.34	26.09	25.09	25.69	27.77	26.27	26.53
TOTAL	**91.80**	**90.77**	**89.31**	**92.41**	**89.37**	**89.39**	**90.47**	**89.91**	**93.20**	**88.77**
RANK	**6.00**	**11.00**	**21.00**	**3.00**	**20.00**	**19.00**	**13.00**	**16.00**	**1.00**	**23.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	88.28	87.78	88.11	87.63	86.96	87.62	85.69	82.85	84.45	84.11
NORMAL. ANY HYP CORRECT	37.75	42.51	53.71	34.75	52.20	45.27	46.77	50.78	33.59	53.72
LANGUAGE IMPROVEMENT	25.70	24.79	25.36	26.16	23.96	26.89	24.18	23.09	24.89	21.37
SPEAKER IMPROVEMENT	31.00	31.67	32.47	32.14	30.84	31.04	30.01	28.23	27.76	28.86
ACTIVITY IMPROVEMENT	-1.78	-2.04	-1.79	-1.44	-2.76	-2.69	-3.25	-3.03	-1.88	-4.37
PLATFORM IMPROVEMENT	22.79	21.90	21.63	20.56	22.05	22.94	20.97	23.01	21.14	21.59
LSP IMPROVEMENT	58.94	58.12	59.11	59.27	57.74	58.71	57.19	54.86	56.48	55.29
ROUTING										
COMPLETELY CORRECT	97.56	97.56	97.43	97.60	97.56	97.43	97.43	97.55	97.81	97.56
PARTIALLY CORRECT	97.20	97.20	97.04	97.25	97.20	97.04	97.04	97.19	97.48	97.20
ACCURACY	97.20	97.20	97.04	97.25	97.20	97.04	97.04	97.19	97.48	97.20
EFFICIENCY	89.21	89.42	87.15	85.77	89.56	88.83	88.13	87.51	81.33	88.97
TOTAL	**643.84**	**646.12**	**657.26**	**636.92**	**652.51**	**650.12**	**641.21**	**639.21**	**620.53**	**641.49**
RANK	**14.00**	**11.00**	**3.00**	**23.00**	**6.00**	**8.00**	**17.00**	**20.00**	**37.00**	**16.00**
GRAND TOTAL	**1286.42**	**1281.54**	**1282.43**	**1283.80**	**1278.12**	**1275.87**	**1274.53**	**1268.62**	**1272.91**	**1262.85**
RANK FOR GRAND TOTAL	**10.00**	**13.00**	**12.00**	**11.00**	**14.00**	**15.00**	**16.00**	**18.00**	**17.00**	**19.00**
TOTAL RANK	**20.00**	**22.00**	**24.00**	**26.00**	**26.00**	**27.00**	**30.00**	**36.00**	**38.00**	**39.00**

SCENARIO GROUP (4,5,6)	100/100/25	75/75/50	75/100/25	100/75/0	100/75/25	100/25/25	100/50/50	75/50/25	100/100/50	100/75/50
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	26.46	26.17	27.97	26.59	28.37	27.16	28.54	28.27	27.77	27.76
SPEAKER IMPROVEMENT	36.23	34.96	35.77	36.52	37.72	35.04	34.35	36.13	34.41	37.03
ACTIVITY IMPROVEMENT	5.68	5.68	5.68	0.00	5.68	5.68	5.68	5.68	5.68	5.68
PLATFORM IMPROVEMENT	19.68	19.30	14.40	20.27	20.12	21.35	22.25	19.91	17.63	20.14
TOTAL	**88.04**	**86.11**	**83.82**	**83.39**	**91.90**	**89.23**	**90.83**	**89.99**	**85.49**	**90.61**
RANK	**15.00**	**23.00**	**28.00**	**31.00**	**2.00**	**12.00**	**5.00**	**10.00**	**25.00**	**7.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	83.10	86.79	89.02	85.85	72.44	82.59	78.04	75.13	81.76	73.13
NORMAL. ANY HYP CORRECT	38.13	44.03	36.68	33.40	36.98	30.86	29.90	34.67	35.71	33.14
LANGUAGE IMPROVEMENT	23.83	23.56	26.60	25.05	25.22	23.87	25.62	25.16	24.78	24.28
SPEAKER IMPROVEMENT	24.94	28.86	30.67	27.87	15.30	24.76	20.61	17.90	24.14	15.78
ACTIVITY IMPROVEMENT	10.16	11.40	13.26	12.16	-1.91	7.50	5.30	-0.03	9.43	-1.68
PLATFORM IMPROVEMENT	21.92	22.54	20.85	25.30	21.34	20.52	20.77	20.93	19.54	21.53
LSP IMPROVEMENT	53.45	56.41	58.64	56.13	43.10	51.91	48.82	45.39	51.80	42.93
ROUTING										
COMPLETELY CORRECT	99.72	98.52	99.50	99.95	99.28	99.29	99.26	99.25	99.19	99.27
PARTIALLY CORRECT	99.68	98.33	99.44	99.95	99.19	99.20	99.16	99.15	99.09	99.18
ACCURACY	99.68	98.33	99.44	99.95	99.19	99.20	99.16	99.15	99.09	99.18
EFFICIENCY	50.10	55.14	56.65	64.15	43.48	33.66	29.83	43.18	51.57	38.45
TOTAL	**604.71**	**623.91**	**630.74**	**629.75**	**553.61**	**573.36**	**556.47**	**559.87**	**596.08**	**545.18**
RANK	**15.00**	**8.00**	**4.00**	**6.00**	**35.00**	**26.00**	**34.00**	**31.00**	**17.00**	**37.00**
GRAND TOTAL	**1221.03**	**1226.70**	**1217.48**	**1213.46**	**1196.89**	**1197.98**	**1192.28**	**1189.79**	**1194.84**	**1179.46**
RANK FOR GRAND TOTAL	**11.00**	**8.00**	**12.00**	**13.00**	**15.00**	**14.00**	**17.00**	**18.00**	**16.00**	**23.00**
TOTAL RANK	**30.00**	**31.00**	**32.00**	**37.00**	**37.00**	**38.00**	**39.00**	**41.00**	**42.00**	**44.00**

SCENARIO GROUP (1,2,3)	25/75/0	100/75/50	100/25/25	75/75/50	100/50/50	50/75/50	75/50/75	100/100/75	100/75/75	75/100/100
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	26.54	27.33	27.99	28.46	28.45	29.04	27.62	27.69	27.87	29.03
SPEAKER IMPROVEMENT	34.82	36.84	36.13	35.48	35.88	32.31	33.40	35.88	34.76	32.11
ACTIVITY IMPROVEMENT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PLATFORM IMPROVEMENT	28.34	25.33	26.67	22.90	24.75	23.65	24.81	23.80	23.40	21.95
TOTAL	**89.70**	**89.50**	**90.78**	**86.85**	**89.09**	**85.00**	**85.82**	**87.37**	**86.03**	**83.09**
RANK	**17.00**	**18.00**	**10.00**	**28.00**	**22.00**	**35.00**	**33.00**	**27.00**	**32.00**	**43.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	82.36	84.16	83.56	84.30	84.76	83.91	84.16	83.23	84.79	83.78
NORMAL. ANY HYP CORRECT	51.83	40.79	35.32	43.93	36.02	55.63	52.83	46.45	46.70	57.99
LANGUAGE IMPROVEMENT	22.44	25.75	24.93	26.40	26.33	24.32	23.32	23.76	24.45	24.61
SPEAKER IMPROVEMENT	27.61	28.70	28.30	29.16	29.57	27.17	27.78	28.44	29.71	27.12
ACTIVITY IMPROVEMENT	-3.89	-3.19	-3.96	-3.32	-3.07	-4.51	-4.56	-3.60	-3.52	-3.99
PLATFORM IMPROVEMENT	23.62	22.27	20.84	19.99	21.89	18.61	20.27	19.31	18.91	17.50
LSP IMPROVEMENT	54.37	55.46	55.47	55.23	56.50	55.14	55.57	54.11	55.09	54.77
ROUTING										
COMPLETELY CORRECT	95.75	97.44	97.57	97.42	98.20	97.55	97.68	97.43	97.68	97.54
PARTIALLY CORRECT	95.11	97.05	97.21	97.03	97.93	97.19	97.33	97.04	97.33	97.18
ACCURACY	95.11	97.05	97.21	97.03	97.93	97.19	97.33	97.04	97.33	97.18
EFFICIENCY	88.33	86.98	82.81	86.93	83.15	87.89	86.57	87.32	84.67	88.57
TOTAL	**632.64**	**632.45**	**619.26**	**634.10**	**629.21**	**640.09**	**638.27**	**630.53**	**633.18**	**642.26**
RANK	**27.00**	**29.00**	**38.00**	**24.00**	**31.00**	**18.00**	**21.00**	**30.00**	**26.00**	**15.00**
GRAND TOTAL	**1260.52**	**1258.99**	**1254.72**	**1242.03**	**1252.83**	**1235.07**	**1239.02**	**1242.12**	**1235.37**	**1223.86**
RANK FOR GRAND TOTAL	**20.00**	**21.00**	**22.00**	**26.00**	**23.00**	**30.00**	**27.00**	**25.00**	**29.00**	**31.00**
TOTAL RANK	**44.00**	**47.00**	**48.00**	**52.00**	**53.00**	**53.00**	**54.00**	**57.00**	**58.00**	**58.00**

SCENARIO GROUP (4,5,6)	75/100/75	100/100/0	100/100/75	100/100/100	75/25/25	50/75/25	100/25/0	75/25/0	100/75/100	100/50/0
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	25.83	26.71	26.63	25.83	27.35	25.69	27.61	27.19	28.13	28.30
SPEAKER IMPROVEMENT	31.59	35.75	36.73	36.23	34.93	34.09	34.43	33.82	36.01	34.31
ACTIVITY IMPROVEMENT	5.68	0.00	5.68	5.68	5.68	5.68	0.00	0.00	5.68	0.00
PLATFORM IMPROVEMENT	16.47	14.87	18.07	17.67	22.76	16.20	21.68	21.58	18.73	17.45
TOTAL	**79.57**	**77.34**	**87.12**	**85.41**	**90.72**	**81.66**	**83.73**	**82.59**	**88.55**	**80.06**
RANK	**40.00**	**43.00**	**20.00**	**26.00**	**6.00**	**37.00**	**29.00**	**35.00**	**13.00**	**39.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	84.66	88.80	73.21	78.18	72.07	82.59	85.11	84.55	65.56	83.57
NORMAL. ANY HYP CORRECT	44.06	36.56	45.55	40.55	31.74	40.48	26.03	29.61	36.08	31.04
LANGUAGE IMPROVEMENT	24.11	25.20	18.50	21.47	22.90	20.85	23.80	25.16	20.50	26.65
SPEAKER IMPROVEMENT	25.60	30.81	15.93	20.14	13.86	24.34	27.32	26.62	7.56	26.48
ACTIVITY IMPROVEMENT	10.35	13.59	-0.56	6.10	-2.67	10.41	11.99	10.78	-7.28	10.62
PLATFORM IMPROVEMENT	21.00	22.21	16.72	16.34	21.68	19.65	23.28	22.70	14.52	21.30
LSP IMPROVEMENT	53.29	57.02	42.09	47.14	42.06	51.77	55.43	54.16	36.03	53.27
ROUTING										
COMPLETELY CORRECT	99.03	99.84	98.59	98.82	99.30	99.02	99.99	99.99	99.11	99.90
PARTIALLY CORRECT	98.91	99.82	98.41	98.67	99.21	98.90	99.99	99.99	99.00	99.89
ACCURACY	98.91	99.82	98.41	98.67	99.21	98.90	99.99	99.99	99.00	99.89
EFFICIENCY	70.35	59.94	62.97	58.95	34.01	64.68	30.32	39.00	51.60	36.84
TOTAL	**630.27**	**633.61**	**569.82**	**585.01**	**533.36**	**611.61**	**583.25**	**592.55**	**521.68**	**589.45**
RANK	**5.00**	**3.00**	**27.00**	**22.00**	**42.00**	**12.00**	**23.00**	**18.00**	**44.00**	**20.00**
GRAND TOTAL	**1187.24**	**1174.97**	**1179.63**	**1182.88**	**1168.36**	**1183.23**	**1169.34**	**1170.69**	**1141.51**	**1149.84**
RANK FOR GRAND TOTAL	**19.00**	**24.00**	**22.00**	**21.00**	**27.00**	**20.00**	**26.00**	**25.00**	**30.00**	**28.00**
TOTAL RANK	**45.00**	**46.00**	**47.00**	**48.00**	**48.00**	**49.00**	**52.00**	**53.00**	**57.00**	**59.00**

SCENARIO GROUP (1,2,3)	75/50/50	75/25/25	50/75/75	100/100/100	100/75/100	50/100/50	25/100/0	75/100/75	100/50/75	25/75/25
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	28.44	26.68	27.51	28.23	28.54	25.83	26.69	27.63	27.46	26.55
SPEAKER IMPROVEMENT	35.18	36.83	31.42	33.32	32.82	32.12	32.77	33.15	35.91	29.24
ACTIVITY IMPROVEMENT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PLATFORM IMPROVEMENT	22.75	24.55	24.41	22.86	23.65	23.16	27.09	23.36	23.05	27.12
TOTAL	**86.36**	**88.05**	**83.34**	**84.41**	**85.01**	**81.11**	**86.55**	**84.14**	**86.43**	**82.91**
RANK	**31.00**	**26.00**	**42.00**	**36.00**	**34.00**	**50.00**	**29.00**	**37.00**	**30.00**	**45.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	85.15	85.28	82.62	81.41	81.93	84.16	77.42	79.88	80.05	82.49
NORMAL. ANY HYP CORRECT	40.75	36.39	56.97	56.95	48.69	55.80	47.37	50.32	36.55	50.39
LANGUAGE IMPROVEMENT	26.07	24.75	22.46	22.16	24.57	22.44	22.13	22.56	24.85	23.30
SPEAKER IMPROVEMENT	29.87	30.07	25.98	25.39	25.71	29.33	21.76	23.52	26.02	26.44
ACTIVITY IMPROVEMENT	-3.33	-3.05	-4.45	-5.99	-6.50	-3.15	-4.93	-4.10	-6.62	-3.09
PLATFORM IMPROVEMENT	19.53	21.02	19.36	16.58	19.80	19.04	21.36	18.94	18.99	22.42
LSP IMPROVEMENT	55.81	55.35	53.75	52.78	52.83	54.25	48.30	50.43	50.80	53.90
ROUTING										
COMPLETELY CORRECT	97.68	97.58	97.54	97.55	97.55	97.43	96.01	97.44	97.68	95.62
PARTIALLY CORRECT	97.33	97.22	97.18	97.19	97.19	97.04	95.41	97.05	97.33	94.96
ACCURACY	97.33	97.22	97.18	97.19	97.19	97.04	95.41	97.05	97.33	94.96
EFFICIENCY	86.31	84.73	89.09	86.54	85.67	86.27	87.36	86.04	82.05	86.22
TOTAL	**632.48**	**626.55**	**637.68**	**627.74**	**624.63**	**639.65**	**607.60**	**619.12**	**606.03**	**627.60**
RANK	**28.00**	**34.00**	**22.00**	**32.00**	**35.00**	**19.00**	**45.00**	**39.00**	**48.00**	**33.00**
GRAND TOTAL	**1237.02**	**1242.92**	**1221.07**	**1218.59**	**1219.69**	**1207.40**	**1213.44**	**1208.14**	**1210.01**	**1207.96**
RANK FOR GRAND TOTAL	**28.00**	**24.00**	**32.00**	**34.00**	**33.00**	**39.00**	**35.00**	**37.00**	**36.00**	**38.00**
TOTAL RANK	**59.00**	**60.00**	**64.00**	**68.00**	**69.00**	**69.00**	**74.00**	**76.00**	**78.00**	**78.00**

SCENARIO GROUP (4,5,6)	50/100/25	75/100/0	75/25/100	75/100/100	75/50/0	100/25/50	75/100/50	50/75/100	50/100/75	50/50/100
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	27.28	25.74	26.57	27.31	28.64	27.06	26.54	25.59	26.67	26.12
SPEAKER IMPROVEMENT	35.10	34.39	35.37	32.26	35.74	33.61	35.00	33.69	32.29	31.52
ACTIVITY IMPROVEMENT	5.68	0.00	5.68	5.68	0.00	5.68	5.68	5.68	5.68	5.68
PLATFORM IMPROVEMENT	11.09	15.22	19.68	16.56	19.83	23.28	15.38	15.32	11.08	20.12
TOTAL	**79.14**	**75.35**	**87.29**	**81.80**	**84.21**	**89.63**	**82.60**	**80.28**	**75.72**	**83.44**
RANK	**41.00**	**47.00**	**17.00**	**36.00**	**27.00**	**11.00**	**34.00**	**38.00**	**46.00**	**30.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	78.71	84.21	70.33	72.41	69.74	64.58	73.32	74.45	76.35	69.73
NORMAL. ANY HYP CORRECT	39.08	35.21	32.00	42.20	33.53	27.04	42.82	34.54	40.60	32.87
LANGUAGE IMPROVEMENT	23.18	24.37	22.38	22.60	24.73	22.10	22.73	22.32	24.97	22.09
SPEAKER IMPROVEMENT	21.12	26.49	14.46	14.96	12.78	7.24	16.19	17.15	18.83	12.77
ACTIVITY IMPROVEMENT	8.81	10.15	-5.09	-1.55	-5.73	-9.90	-0.59	6.41	7.86	0.04
PLATFORM IMPROVEMENT	16.84	22.36	12.88	18.71	21.15	16.69	19.97	17.16	16.10	16.53
LSP IMPROVEMENT	48.33	53.77	39.07	43.31	41.11	33.32	43.19	44.02	45.92	40.12
ROUTING										
COMPLETELY CORRECT	99.07	99.56	99.11	98.76	99.98	99.29	98.46	98.80	98.67	98.88
PARTIALLY CORRECT	98.95	99.51	99.00	98.60	99.98	99.20	98.26	98.64	98.50	98.73
ACCURACY	98.95	99.51	99.00	98.60	99.98	99.20	98.26	98.64	98.50	98.73
EFFICIENCY	56.42	49.72	35.12	57.99	45.41	23.35	46.48	51.56	55.29	45.48
TOTAL	**589.46**	**604.84**	**518.25**	**566.58**	**542.64**	**482.10**	**559.10**	**563.70**	**581.58**	**535.96**
RANK	**19.00**	**14.00**	**45.00**	**28.00**	**38.00**	**55.00**	**33.00**	**30.00**	**24.00**	**41.00**
GRAND TOTAL	**1143.44**	**1132.29**	**1129.29**	**1139.21**	**1132.10**	**1109.51**	**1137.27**	**1125.65**	**1111.59**	**1120.04**
RANK FOR GRAND TOTAL	**29.00**	**33.00**	**35.00**	**31.00**	**34.00**	**41.00**	**32.00**	**36.00**	**40.00**	**37.00**
TOTAL RANK	**60.00**	**61.00**	**62.00**	**64.00**	**65.00**	**66.00**	**67.00**	**68.00**	**70.00**	**71.00**

SCENARIO GROUP (1,2,3)	100/50/100	50/100/100	75/25/50	100/25/50	75/50/100	50/75/100	50/50/75	100/25/75	75/75/100	50/100/75
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	27.18	26.20	28.10	27.28	28.42	27.04	27.44	28.16	27.56	26.36
SPEAKER IMPROVEMENT	34.24	31.38	35.96	35.85	32.68	30.91	32.00	34.94	30.95	29.29
ACTIVITY IMPROVEMENT	-0.13	0.00	-0.13	-0.13	-0.13	0.00	0.00	-0.13	0.00	0.00
PLATFORM IMPROVEMENT	22.43	22.20	24.45	25.63	22.42	22.78	21.97	20.79	21.88	23.42
TOTAL	**83.71**	**79.78**	**88.38**	**88.63**	**83.39**	**80.72**	**81.41**	**83.75**	**80.38**	**79.07**
RANK	**39.00**	**55.00**	**25.00**	**24.00**	**41.00**	**51.00**	**49.00**	**38.00**	**53.00**	**56.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	81.76	81.67	75.81	79.03	78.04	80.17	80.31	79.63	79.36	78.46
NORMAL. ANY HYP CORRECT	44.33	57.91	30.01	25.70	41.65	49.50	48.30	37.43	49.40	54.25
LANGUAGE IMPROVEMENT	24.06	22.17	26.31	25.73	25.72	23.22	23.77	26.03	23.92	21.18
SPEAKER IMPROVEMENT	26.13	26.89	23.55	24.31	22.39	24.08	23.44	25.92	23.20	21.35
ACTIVITY IMPROVEMENT	-6.72	-3.88	-10.12	-7.48	-8.02	-5.38	-5.47	-8.80	-6.75	-5.00
PLATFORM IMPROVEMENT	18.62	17.49	18.78	20.92	19.40	19.72	18.83	17.36	17.63	17.85
LSP IMPROVEMENT	52.09	52.37	48.26	50.11	49.05	50.92	50.19	50.35	49.96	48.60
ROUTING										
COMPLETELY CORRECT	97.54	97.55	98.06	98.19	98.12	97.44	97.67	97.68	97.55	97.43
PARTIALLY CORRECT	97.18	97.19	97.77	97.92	97.84	97.05	97.32	97.33	97.19	97.04
ACCURACY	97.18	97.19	97.77	97.92	97.84	97.05	97.32	97.33	97.19	97.04
EFFICIENCY	84.68	87.19	68.45	59.45	84.05	87.47	87.25	77.18	86.56	87.08
TOTAL	**616.85**	**633.73**	**574.65**	**571.81**	**606.08**	**621.24**	**618.95**	**597.44**	**615.21**	**615.28**
RANK	**41.00**	**25.00**	**58.00**	**60.00**	**46.00**	**36.00**	**40.00**	**53.00**	**43.00**	**42.00**
GRAND TOTAL	**1202.84**	**1192.22**	**1193.30**	**1192.20**	**1189.81**	**1186.26**	**1188.82**	**1183.70**	**1177.89**	**1168.75**
RANK FOR GRAND TOTAL	**40.00**	**42.00**	**41.00**	**43.00**	**44.00**	**46.00**	**45.00**	**47.00**	**48.00**	**49.00**
TOTAL RANK	**80.00**	**80.00**	**83.00**	**84.00**	**87.00**	**87.00**	**89.00**	**91.00**	**96.00**	**98.00**

SCENARIO GROUP (4,5,6)	75/50/100	50/75/50	75/75/100	50/100/0	50/75/75	100/25/100	50/25/100	25/75/75	25/75/50	50/100/50
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	27.22	27.18	28.35	25.77	26.10	27.30	27.37	25.71	24.18	26.12
SPEAKER IMPROVEMENT	33.31	33.72	33.43	36.18	33.89	33.15	33.24	27.08	30.80	35.05
ACTIVITY IMPROVEMENT	5.68	5.68	5.68	0.00	5.68	5.68	5.68	5.68	5.68	5.68
PLATFORM IMPROVEMENT	20.15	16.13	15.76	9.41	11.48	22.06	19.70	3.20	2.01	10.93
TOTAL	**86.35**	**82.71**	**83.22**	**71.36**	**77.15**	**88.19**	**85.99**	**61.68**	**62.67**	**77.78**
RANK	**22.00**	**33.00**	**32.00**	**50.00**	**44.00**	**14.00**	**24.00**	**61.00**	**60.00**	**42.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	67.21	65.20	66.12	78.62	71.28	33.82	51.58	76.12	72.38	59.88
NORMAL. ANY HYP CORRECT	33.07	39.85	35.96	32.80	42.63	29.28	27.03	35.81	37.50	37.11
LANGUAGE IMPROVEMENT	22.34	18.45	23.10	24.65	21.86	13.49	18.59	22.34	23.05	16.14
SPEAKER IMPROVEMENT	10.44	7.20	10.04	21.58	15.01	-20.72	-4.39	18.78	15.71	2.53
ACTIVITY IMPROVEMENT	-6.57	-4.47	-5.72	7.53	-1.44	-39.05	-21.35	6.59	9.33	-7.48
PLATFORM IMPROVEMENT	11.40	15.47	14.58	19.89	16.05	3.78	6.42	14.79	14.09	7.54
LSP IMPROVEMENT	36.40	35.74	37.06	49.26	40.62	3.41	21.53	44.83	39.79	28.70
ROUTING										
COMPLETELY CORRECT	98.93	98.91	98.86	99.24	98.26	99.16	99.02	98.34	98.14	98.63
PARTIALLY CORRECT	98.80	98.77	98.71	99.14	98.03	99.05	98.90	98.12	97.90	98.45
ACCURACY	98.80	98.77	98.71	99.14	98.03	99.05	98.90	98.12	97.90	98.45
EFFICIENCY	42.07	63.23	47.68	45.75	52.43	26.28	28.02	51.88	53.48	54.53
TOTAL	**512.89**	**537.12**	**525.09**	**577.59**	**552.77**	**347.56**	**424.26**	**565.71**	**559.26**	**494.43**
RANK	**49.00**	**40.00**	**43.00**	**25.00**	**36.00**	**66.00**	**62.00**	**29.00**	**32.00**	**52.00**
GRAND TOTAL	**1117.37**	**1116.06**	**1107.60**	**1077.10**	**1092.78**	**964.89**	**1026.22**	**997.43**	**997.94**	**1038.96**
RANK FOR GRAND TOTAL	**38.00**	**39.00**	**42.00**	**44.00**	**43.00**	**55.00**	**46.00**	**51.00**	**50.00**	**45.00**
TOTAL RANK	**71.00**	**73.00**	**75.00**	**75.00**	**80.00**	**80.00**	**86.00**	**90.00**	**92.00**	**94.00**

SCENARIO GROUP (1,2,3)	75/25/100	25/100/25	25/75/75	50/25/75	25/100/100	50/50/100	100/25/100	25/75/50	75/25/75	25/100/50
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	27.24	25.98	26.57	28.53	26.61	26.68	27.32	25.93	28.28	26.69
SPEAKER IMPROVEMENT	33.68	29.30	29.72	33.15	28.09	30.44	34.05	28.37	32.72	27.65
ACTIVITY IMPROVEMENT	-0.13	0.00	0.00	-0.13	0.00	-0.13	-0.13	0.00	-0.13	0.00
PLATFORM IMPROVEMENT	21.61	26.62	23.84	21.44	21.87	20.64	22.40	22.32	21.36	23.71
TOTAL	**82.39**	**81.90**	**80.13**	**83.00**	**76.58**	**77.64**	**83.64**	**76.62**	**82.23**	**78.04**
RANK	**46.00**	**48.00**	**54.00**	**44.00**	**61.00**	**59.00**	**40.00**	**60.00**	**47.00**	**57.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	81.24	73.75	74.73	75.64	79.13	78.44	50.07	78.10	72.92	71.61
NORMAL. ANY HYP CORRECT	30.29	46.46	51.78	28.83	51.51	44.83	21.88	54.35	28.09	47.34
LANGUAGE IMPROVEMENT	26.17	21.11	21.70	26.86	22.53	23.24	23.44	20.07	25.95	22.55
SPEAKER IMPROVEMENT	26.74	17.18	18.54	21.91	22.97	21.53	-1.05	22.39	19.11	15.66
ACTIVITY IMPROVEMENT	-7.83	-5.16	-4.51	-12.35	-3.97	-7.03	-35.02	-6.50	-14.25	-4.32
PLATFORM IMPROVEMENT	19.53	20.24	18.20	17.96	17.18	17.03	14.18	15.04	17.19	18.49
LSP IMPROVEMENT	52.03	44.09	46.15	46.70	49.82	47.55	21.94	48.45	43.17	42.11
ROUTING										
COMPLETELY CORRECT	97.68	95.75	97.20	97.95	95.62	98.07	98.58	95.75	97.96	95.62
PARTIALLY CORRECT	97.33	95.11	96.77	97.64	94.96	97.78	98.37	95.11	97.65	94.96
ACCURACY	97.33	95.11	96.77	97.64	94.96	97.78	98.37	95.11	97.65	94.96
EFFICIENCY	70.71	85.79	86.43	70.89	86.83	86.14	42.46	86.80	68.11	86.60
TOTAL	**591.21**	**589.43**	**603.76**	**569.66**	**611.52**	**605.35**	**433.22**	**604.68**	**553.55**	**585.58**
RANK	**54.00**	**55.00**	**50.00**	**61.00**	**44.00**	**47.00**	**66.00**	**49.00**	**65.00**	**56.00**
GRAND TOTAL	**1167.93**	**1162.74**	**1164.64**	**1150.64**	**1147.55**	**1148.81**	**1018.71**	**1141.01**	**1129.17**	**1131.89**
RANK FOR GRAND TOTAL	50.00	52.00	51.00	53.00	55.00	54.00	66.00	56.00	60.00	59.00
TOTAL RANK	100.00	103.00	104.00	105.00	105.00	106.00	106.00	109.00	112.00	113.00

SCENARIO GROUP (4,5,6)	50/100/100	25/75/25	50/75/0	25/50/75	25/75/0	25/25/100	25/100/75	100/25/75	25/75/100	75/25/75
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	27.19	26.83	26.60	27.38	26.45	27.33	26.89	27.23	27.64	29.24
SPEAKER IMPROVEMENT	31.05	34.16	36.34	25.87	32.54	28.23	29.88	34.53	27.35	35.12
ACTIVITY IMPROVEMENT	5.68	5.68	0.00	5.68	0.00	5.68	5.68	-16.04	5.68	-16.04
PLATFORM IMPROVEMENT	9.08	3.07	10.56	7.84	-2.67	15.80	5.05	21.39	3.48	20.85
TOTAL	**73.00**	**69.74**	**73.50**	**66.77**	**56.31**	**77.04**	**67.50**	**67.11**	**64.15**	**69.17**
RANK	**49.00**	**51.00**	**48.00**	**56.00**	**65.00**	**45.00**	**54.00**	**55.00**	**57.00**	**52.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	64.32	65.15	63.72	66.45	72.04	54.37	60.61	58.46	56.42	51.11
NORMAL. ANY HYP CORRECT	37.17	37.38	32.54	32.84	29.98	48.14	28.91	33.01	30.49	32.16
LANGUAGE IMPROVEMENT	23.93	21.34	23.09	24.20	25.87	-8.28	26.10	18.60	23.51	17.12
SPEAKER IMPROVEMENT	7.71	6.70	6.63	8.77	13.54	-3.61	2.38	2.37	-1.58	-4.91
ACTIVITY IMPROVEMENT	-5.41	-5.54	-7.17	-5.35	6.80	-18.73	0.84	-18.06	-5.86	-21.86
PLATFORM IMPROVEMENT	12.58	9.95	17.35	15.54	14.73	-12.05	11.26	11.05	13.75	4.53
LSP IMPROVEMENT	33.53	34.35	34.19	35.73	41.36	24.15	30.33	28.81	26.18	21.43
ROUTING										
COMPLETELY CORRECT	98.65	98.07	99.43	98.88	98.79	98.31	98.57	99.20	98.68	99.12
PARTIALLY CORRECT	98.47	97.82	99.36	98.73	98.63	98.10	98.38	99.10	98.51	99.01
ACCURACY	98.47	97.82	99.36	98.73	98.63	98.10	98.38	99.10	98.51	99.01
EFFICIENCY	44.42	52.47	40.03	40.31	37.95	77.61	36.71	33.72	33.22	33.90
TOTAL	**513.86**	**515.51**	**508.54**	**514.83**	**538.33**	**456.11**	**492.47**	**465.35**	**471.83**	**430.62**
RANK	**48.00**	**46.00**	**51.00**	**47.00**	**39.00**	**60.00**	**53.00**	**58.00**	**56.00**	**61.00**
GRAND TOTAL	1024.86	1003.70	1023.04	982.20	932.50	995.40	964.95	935.11	920.87	914.77
RANK FOR GRAND TOTAL	47.00	49.00	48.00	53.00	57.00	52.00	54.00	56.00	58.00	59.00
TOTAL RANK	97.00	97.00	99.00	103.00	104.00	105.00	107.00	113.00	113.00	113.00

SCENARIO GROUP (1,2,3)	50/25/100	25/50/75	25/50/100	25/25/75	25/25/100	25/75/100	25/100/75
MERGE ALGORITHM							
LANGUAGE IMPROVEMENT	28.77	25.72	26.10	26.66	27.42	26.39	27.23
SPEAKER IMPROVEMENT	30.87	28.30	25.73	30.26	26.78	27.54	27.77
ACTIVITY IMPROVEMENT	-0.13	0.00	-0.13	-0.13	-0.13	0.00	0.00
PLATFORM IMPROVEMENT	21.05	22.43	21.19	21.05	22.19	21.69	21.46
TOTAL	**80.56**	**76.45**	**72.89**	**77.85**	**76.26**	**75.61**	**76.46**
RANK	**52.00**	**63.00**	**66.00**	**58.00**	**64.00**	**65.00**	**62.00**
CORRELATION ALGORITHM							
TOP HYPOTHESIS CORRECT	75.76	75.92	77.09	75.09	73.38	71.80	70.36
NORMAL. ANY HYP CORRECT	27.15	46.32	47.37	30.09	62.20	45.20	47.42
LANGUAGE IMPROVEMENT	27.83	21.65	22.89	25.06	12.40	21.89	19.95
SPEAKER IMPROVEMENT	20.24	20.65	19.72	17.81	15.62	15.39	13.35
ACTIVITY IMPROVEMENT	-12.22	-6.65	-6.40	-13.31	-15.49	-9.41	-7.93
PLATFORM IMPROVEMENT	19.10	17.81	16.98	19.50	6.60	15.66	12.59
LSP IMPROVEMENT	46.66	46.61	47.03	45.01	44.05	41.74	40.62
ROUTING							
COMPLETELY CORRECT	98.07	97.50	97.30	98.25	97.39	95.62	95.60
PARTIALLY CORRECT	97.78	97.13	96.89	97.98	97.00	94.96	94.94
ACCURACY	97.78	97.13	96.89	97.98	97.00	94.96	94.94
EFFICIENCY	70.60	86.36	86.75	75.07	85.76	85.05	85.80
TOTAL	**568.75**	**600.44**	**602.52**	**568.52**	**575.90**	**572.87**	**567.64**
RANK	**62.00**	**52.00**	**51.00**	**63.00**	**57.00**	**59.00**	**64.00**
GRAND TOTAL	**1132.68**	**1135.60**	**1112.73**	**1113.45**	**1109.69**	**1102.17**	**1102.86**
RANK FOR GRAND TOTAL	**58.00**	**57.00**	**62.00**	**61.00**	**63.00**	**65.00**	**64.00**
TOTAL RANK	**114.00**	**115.00**	**117.00**	**121.00**	**121.00**	**124.00**	**126.00**

SCENARIO GROUP (4,5,6)	25/100/50	25/100/0	25/100/25	25/50/100	25/100/100	50/25/75	25/25/75
MERGE ALGORITHM							
LANGUAGE IMPROVEMENT	25.20	27.02	26.31	28.47	24.79	26.29	26.47
SPEAKER IMPROVEMENT	26.89	32.11	28.00	25.16	24.86	33.00	30.93
ACTIVITY IMPROVEMENT	5.68	0.00	5.68	5.68	5.68	-16.04	-16.04
PLATFORM IMPROVEMENT	3.04	-4.21	2.89	9.04	4.38	20.64	17.00
TOTAL	**60.80**	**54.92**	**62.88**	**68.35**	**59.71**	**63.89**	**58.35**
RANK	**62.00**	**66.00**	**59.00**	**53.00**	**63.00**	**58.00**	**64.00**
CORRELATION ALGORITHM							
TOP HYPOTHESIS CORRECT	60.25	71.70	56.38	45.44	58.93	49.05	50.60
NORMAL. ANY HYP CORRECT	31.01	27.48	30.17	40.06	28.90	30.50	33.41
LANGUAGE IMPROVEMENT	20.34	25.24	24.24	2.85	21.63	17.64	1.37
SPEAKER IMPROVEMENT	3.13	13.61	-1.40	-12.10	0.28	-7.99	-6.41
ACTIVITY IMPROVEMENT	-5.10	1.53	-7.63	-21.36	-3.02	-21.65	-20.19
PLATFORM IMPROVEMENT	11.61	5.81	10.05	-1.72	6.44	9.07	1.42
LSP IMPROVEMENT	29.30	40.87	25.38	15.73	27.19	17.79	19.78
ROUTING							
COMPLETELY CORRECT	98.04	98.83	97.59	98.30	98.63	99.02	99.02
PARTIALLY CORRECT	97.79	98.68	97.28	98.08	98.45	98.90	98.90
ACCURACY	97.79	98.68	97.28	98.08	98.45	98.90	98.90
EFFICIENCY	43.83	30.31	33.26	44.46	34.88	30.27	41.84
TOTAL	**488.00**	**512.74**	**462.60**	**407.82**	**470.77**	**421.51**	**418.63**
RANK	**54.00**	**50.00**	**59.00**	**65.00**	**57.00**	**63.00**	**64.00**
GRAND TOTAL	**913.63**	**897.18**	**902.75**	**886.29**	**888.76**	**868.75**	**827.09**
RANK FOR GRAND TOTAL	**60.00**	**62.00**	**61.00**	**64.00**	**63.00**	**65.00**	**66.00**
TOTAL RANK	**116.00**	**116.00**	**118.00**	**118.00**	**120.00**	**121.00**	**128.00**

SCENARIO GROUP (7,8,9)	75/50/0	75/75/25	100/100/0	75/25/0	100/75/0	100/50/25	100/50/0	75/50/25	50/75/25
MERGE ALGORITHM									
LANGUAGE IMPROVEMENT	27.18	28.03	26.89	27.45	25.29	27.02	27.16	27.61	26.77
SPEAKER IMPROVEMENT	36.36	35.07	32.57	35.30	34.84	36.08	33.55	32.70	33.51
ACTIVITY IMPROVEMENT	0.00	-1.47	0.00	0.00	0.00	-0.49	0.00	-0.49	-1.47
PLATFORM IMPROVEMENT	26.24	26.42	25.86	27.19	26.51	26.47	26.10	24.34	25.47
TOTAL	**89.78**	**88.05**	**85.32**	**89.94**	**86.64**	**89.08**	**86.81**	**84.15**	**84.27**
RANK	**3.00**	**5.00**	**12.00**	**1.00**	**10.00**	**4.00**	**9.00**	**17.00**	**16.00**
CORRELATION ALGORITHM									
TOP HYPOTHESIS CORRECT	66.56	62.19	70.01	62.33	63.95	63.46	61.87	66.04	62.60
NORMAL. ANY HYP CORRECT	31.24	35.45	34.07	28.48	32.81	32.64	29.86	32.09	34.89
LANGUAGE IMPROVEMENT	22.89	24.56	23.96	20.92	20.90	21.86	21.44	23.77	22.36
SPEAKER IMPROVEMENT	29.12	24.67	30.37	21.77	25.34	21.87	22.81	26.28	24.19
ACTIVITY IMPROVEMENT	-3.69	-6.40	-2.47	-5.60	-3.97	-7.49	-4.89	-4.54	-4.94
PLATFORM IMPROVEMENT	22.54	22.51	23.77	21.34	22.17	21.74	21.28	21.35	22.54
LSP IMPROVEMENT	55.08	52.03	57.17	48.64	51.48	49.47	49.58	52.84	50.43
ROUTING									
COMPLETELY CORRECT	99.94	96.36	98.39	99.99	99.84	96.79	99.95	96.01	96.81
PARTIALLY CORRECT	99.93	95.77	98.14	99.99	99.82	96.27	99.94	95.36	96.30
ACCURACY	99.93	95.77	98.14	99.99	99.82	96.27	99.94	95.36	96.30
EFFICIENCY	64.19	64.10	64.05	61.72	64.28	63.30	63.07	63.63	64.32
TOTAL	**587.72**	**567.02**	**595.60**	**559.55**	**576.43**	**556.20**	**564.85**	**568.19**	**565.79**
RANK	**2.00**	**6.00**	**1.00**	**12.00**	**4.00**	**13.00**	**8.00**	**5.00**	**7.00**
GRAND TOTAL	**1216.16**	**1183.40**	**1192.81**	**1189.11**	**1182.93**	**1179.76**	**1172.53**	**1157.27**	**1155.66**
RANK FOR GRAND TOTAL	**1.00**	**4.00**	**2.00**	**3.00**	**5.00**	**6.00**	**7.00**	**11.00**	**12.00**
TOTAL RANK	**5.00**	**11.00**	**13.00**	**13.00**	**14.00**	**17.00**	**17.00**	**22.00**	**23.00**

REPRODUCIBLE (1,2,3)	100/75/0	75/50/0	100/100/25	75/75/25	75/100/25	100/50/0	75/50/25	50/75/25	100/50/25
MERGE ALGORITHM									
LANGUAGE IMPROVEMENT	28.32	28.32	28.19	28.19	28.19	28.32	28.19	28.19	28.19
SPEAKER IMPROVEMENT	37.71	37.20	36.17	36.94	35.78	37.07	36.55	35.65	37.32
ACTIVITY IMPROVEMENT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PLATFORM IMPROVEMENT	26.51	26.51	25.87	25.87	25.87	26.51	25.10	25.61	25.22
TOTAL	**92.54**	**92.03**	**90.23**	**91.00**	**89.84**	**91.90**	**89.84**	**89.45**	**90.73**
RANK	**1.00**	**2.00**	**11.00**	**6.00**	**14.00**	**3.00**	**14.00**	**18.00**	**9.00**
CORRELATION ALGORITHM									
TOP HYPOTHESIS CORRECT	88.03	88.80	89.06	87.26	90.99	89.58	88.80	88.29	87.26
NORMAL. ANY HYP CORRECT	44.37	41.70	52.11	50.47	56.37	38.92	45.81	55.09	43.34
LANGUAGE IMPROVEMENT	25.10	25.48	25.36	24.84	26.13	26.38	26.64	24.97	26.64
SPEAKER IMPROVEMENT	30.76	31.79	31.92	29.99	33.98	31.92	32.18	31.79	30.89
ACTIVITY IMPROVEMENT	-2.44	-1.93	-2.19	-2.70	-1.80	-1.54	-2.44	-2.44	-2.19
PLATFORM IMPROVEMENT	22.39	22.39	22.91	22.01	23.29	21.75	23.42	22.39	22.39
LSP IMPROVEMENT	58.17	59.07	59.33	57.14	61.26	59.33	59.07	58.55	57.40
ROUTING									
COMPLETELY CORRECT	97.55	97.55	97.43	97.55	97.43	97.55	97.43	97.43	97.55
PARTIALLY CORRECT	97.19	97.19	97.04	97.19	97.04	97.19	97.04	97.04	97.19
ACCURACY	97.19	97.19	97.04	97.19	97.04	97.19	97.04	97.04	97.19
EFFICIENCY	88.89	88.65	88.27	88.05	89.24	88.53	87.92	88.51	87.00
TOTAL	**647.20**	**647.88**	**658.28**	**648.99**	**670.97**	**646.80**	**652.91**	**658.66**	**644.66**
RANK	**11.00**	**10.00**	**3.00**	**8.00**	**1.00**	**12.00**	**5.00**	**2.00**	**13.00**
GRAND TOTAL	**1294.98**	**1292.09**	**1289.89**	**1285.99**	**1299.85**	**1290.10**	**1281.79**	**1284.81**	**1279.77**
RANK FOR GRAND TOTAL	**2.00**	**3.00**	**5.00**	**6.00**	**1.00**	**4.00**	**8.00**	**7.00**	**11.00**
TOTAL RANK	**12.00**	**12.00**	**14.00**	**14.00**	**15.00**	**15.00**	**19.00**	**20.00**	**22.00**

SCENARIO GROUP (7,8,9)	75/100/0	100/75/25	100/100/50	100/25/0	100/75/50	75/75/50	100/100/25	50/100/0	100/75/75	75/100/25
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	25.95	27.44	27.13	28.04	27.98	26.79	27.34	23.96	27.54	26.96
SPEAKER IMPROVEMENT	31.75	32.80	34.44	35.89	33.87	30.98	32.07	33.71	30.58	32.37
ACTIVITY IMPROVEMENT	0.00	-1.47	-0.49	0.00	-0.24	-0.24	-1.47	0.00	-0.24	-1.47
PLATFORM IMPROVEMENT	25.83	28.06	24.51	25.92	26.07	25.93	25.70	24.36	26.56	25.23
TOTAL	**83.54**	**86.83**	**85.58**	**89.85**	**87.67**	**83.46**	**83.64**	**82.04**	**84.43**	**83.09**
RANK	**21.00**	**8.00**	**11.00**	**2.00**	**6.00**	**22.00**	**20.00**	**28.00**	**14.00**	**25.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	65.66	57.82	60.98	57.67	58.35	64.62	60.91	62.51	59.48	59.02
NORMAL. ANY HYP CORRECT	36.00	34.04	34.03	25.85	31.90	33.32	35.03	34.33	31.76	35.43
LANGUAGE IMPROVEMENT	22.99	22.30	21.76	19.99	21.90	23.43	22.85	20.35	21.39	21.80
SPEAKER IMPROVEMENT	25.21	19.76	22.45	16.53	15.56	21.77	21.62	25.81	15.93	20.85
ACTIVITY IMPROVEMENT	-3.28	-5.40	-5.60	-6.34	-5.85	-3.73	-5.89	-4.66	-6.03	-6.46
PLATFORM IMPROVEMENT	23.58	23.11	19.91	19.32	20.42	22.44	21.07	21.94	21.51	20.97
LSP IMPROVEMENT	52.33	47.35	48.28	43.96	42.67	48.41	48.59	49.11	42.72	47.21
ROUTING										
COMPLETELY CORRECT	98.16	97.40	96.89	99.98	97.02	96.77	97.88	97.49	96.49	97.78
PARTIALLY CORRECT	97.86	96.98	96.38	99.97	96.54	96.26	97.54	97.08	95.93	97.42
ACCURACY	97.86	96.98	96.38	99.97	96.54	96.26	97.54	97.08	95.93	97.42
EFFICIENCY	64.15	63.25	63.40	58.47	62.55	64.49	64.28	62.30	60.74	63.82
TOTAL	**580.53**	**553.60**	**554.87**	**535.37**	**537.60**	**564.04**	**561.40**	**563.33**	**535.85**	**555.26**
RANK	3.00	17.00	15.00	25.00	23.00	9.00	11.00	10.00	24.00	14.00
GRAND TOTAL	**1165.30**	**1161.39**	**1153.97**	**1164.32**	**1151.32**	**1148.28**	**1146.87**	**1137.57**	**1126.87**	**1136.90**
RANK FOR GRAND TOTAL	8.00	10.00	13.00	9.00	14.00	15.00	16.00	17.00	20.00	18.00
TOTAL RANK	24.00	25.00	26.00	27.00	29.00	31.00	31.00	38.00	38.00	39.00

REPRODUCIBLE (1,2,3)	100/75/25	50/75/0	100/100/50	100/100/0	75/100/0	50/100/0	75/25/25	75/25/0	75/100/50	50/100/25
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	28.19	27.93	28.19	28.19	28.19	26.51	28.19	28.32	27.93	26.51
SPEAKER IMPROVEMENT	36.94	35.14	36.42	36.30	35.65	36.17	37.32	37.71	33.85	35.14
ACTIVITY IMPROVEMENT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PLATFORM IMPROVEMENT	25.61	26.38	25.22	26.51	26.38	26.38	24.84	25.87	24.97	25.87
TOTAL	**90.74**	**89.45**	**89.83**	**91.00**	**90.22**	**89.06**	**90.35**	**91.90**	**86.75**	**87.52**
RANK	**8.00**	**18.00**	**17.00**	**6.00**	**12.00**	**20.00**	**10.00**	**3.00**	**28.00**	**24.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	86.87	87.52	88.55	86.74	86.23	85.71	88.03	87.64	86.36	84.68
NORMAL. ANY HYP CORRECT	47.02	52.31	45.63	44.25	49.95	55.94	37.56	35.90	51.76	51.83
LANGUAGE IMPROVEMENT	25.48	24.58	26.77	25.10	23.81	23.17	26.77	25.61	25.61	23.94
SPEAKER IMPROVEMENT	30.12	30.25	31.40	29.35	28.70	29.86	31.92	30.25	29.09	28.70
ACTIVITY IMPROVEMENT	-2.31	-2.57	-2.83	-2.70	-3.34	-3.22	-2.31	-1.41	-2.06	-2.57
PLATFORM IMPROVEMENT	21.62	22.52	23.55	22.26	21.62	23.04	22.14	19.95	21.49	22.65
LSP IMPROVEMENT	57.01	57.40	58.55	56.62	55.98	55.59	58.30	57.52	56.11	54.56
ROUTING										
COMPLETELY CORRECT	97.43	97.55	97.43	97.55	97.55	97.55	97.55	97.55	97.55	97.55
PARTIALLY CORRECT	97.04	97.19	97.04	97.19	97.19	97.19	97.19	97.19	97.19	97.19
ACCURACY	97.04	97.19	97.04	97.19	97.19	97.19	97.19	97.19	97.19	97.19
EFFICIENCY	86.99	89.01	88.27	89.25	89.01	88.77	86.20	85.75	88.41	87.70
TOTAL	**644.31**	**652.95**	**651.40**	**642.80**	**643.89**	**650.79**	**640.54**	**633.14**	**648.70**	**643.42**
RANK	14.00	4.00	6.00	17.00	15.00	7.00	20.00	28.00	9.00	16.00
GRAND TOTAL	**1279.49**	**1279.10**	**1280.21**	**1279.80**	**1275.43**	**1274.21**	**1272.99**	**1276.44**	**1255.95**	**1256.06**
RANK FOR GRAND TOTAL	12.00	13.00	9.00	10.00	15.00	16.00	17.00	14.00	23.00	22.00
TOTAL RANK	22.00	22.00	23.00	23.00	27.00	27.00	30.00	31.00	37.00	40.00

SCENARIO GROUP (7,8,9)	100/100/75	75/50/50	75/100/50	50/75/0	75/25/25	25/75/0	50/100/25	100/50/50	100/25/25	75/25/50
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	27.90	27.77	27.19	26.70	28.09	23.80	25.42	27.76	27.38	28.16
SPEAKER IMPROVEMENT	31.76	31.68	29.56	29.93	35.07	28.95	31.89	32.14	31.41	29.83
ACTIVITY IMPROVEMENT	-0.24	-0.24	-0.49	0.00	-0.24	0.00	-1.47	-0.24	-0.24	-0.24
PLATFORM IMPROVEMENT	24.25	23.68	24.84	23.81	24.69	23.52	26.84	25.10	25.80	25.59
TOTAL	**83.67**	**82.89**	**81.10**	**80.44**	**87.62**	**76.27**	**82.67**	**84.76**	**84.35**	**83.34**
RANK	**18.00**	**26.00**	**29.00**	**30.00**	**7.00**	**36.00**	**27.00**	**13.00**	**15.00**	**24.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	60.91	60.65	59.25	58.61	48.93	59.19	51.40	51.12	52.65	60.39
NORMAL. ANY HYP CORRECT	31.78	31.47	32.95	34.26	29.29	34.04	32.98	29.95	28.36	27.89
LANGUAGE IMPROVEMENT	22.22	21.27	22.57	21.45	20.35	21.12	20.76	20.63	17.92	21.75
SPEAKER IMPROVEMENT	18.85	20.38	20.49	20.55	3.79	22.60	14.61	7.14	9.74	14.57
ACTIVITY IMPROVEMENT	-5.65	-5.75	-5.25	-5.87	-15.10	-4.57	-6.84	-14.34	-11.92	-10.48
PLATFORM IMPROVEMENT	19.22	17.91	20.97	19.85	16.95	20.95	23.55	17.91	16.30	20.64
LSP IMPROVEMENT	45.55	45.95	46.29	45.75	30.13	47.16	40.82	33.87	36.45	41.06
ROUTING										
COMPLETELY CORRECT	96.52	96.79	95.75	98.10	96.80	97.16	96.76	96.80	96.81	96.67
PARTIALLY CORRECT	95.95	96.27	95.06	97.79	96.29	96.70	96.24	96.29	96.30	96.13
ACCURACY	95.95	96.27	95.06	97.79	96.29	96.70	96.24	96.29	96.30	96.13
EFFICIENCY	63.61	63.06	64.04	63.18	58.45	62.68	63.57	59.69	58.80	53.08
TOTAL	**544.91**	**544.25**	**547.19**	**551.48**	**482.17**	**553.71**	**530.09**	**495.34**	**497.72**	**517.83**
RANK	**21.00**	**22.00**	**19.00**	**18.00**	**43.00**	**16.00**	**27.00**	**41.00**	**40.00**	**32.00**
GRAND TOTAL	**1130.62**	**1124.48**	**1114.90**	**1114.58**	**1095.47**	**1087.63**	**1108.79**	**1088.63**	**1088.13**	**1101.21**
RANK FOR GRAND TOTAL	**19.00**	**21.00**	**22.00**	**23.00**	**26.00**	**30.00**	**24.00**	**28.00**	**29.00**	**25.00**
TOTAL RANK	**39.00**	**48.00**	**48.00**	**48.00**	**50.00**	**52.00**	**54.00**	**54.00**	**55.00**	**56.00**

REPRODUCIBLE (1,2,3)	100/25/25	100/75/50	100/25/0	75/75/50	75/50/50	100/50/50	100/100/75	50/75/50	25/75/0	100/75/75
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	28.19	28.19	28.44	28.19	28.19	28.19	28.19	27.93	26.51	28.19
SPEAKER IMPROVEMENT	36.94	36.81	37.84	35.39	35.01	35.78	34.88	31.53	33.98	34.24
ACTIVITY IMPROVEMENT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PLATFORM IMPROVEMENT	24.71	25.10	24.84	24.71	24.32	24.58	24.45	23.55	26.38	23.29
TOTAL	**89.84**	**90.10**	**91.12**	**88.29**	**87.52**	**88.55**	**87.52**	**83.01**	**86.87**	**85.72**
RANK	**14.00**	**13.00**	**5.00**	**23.00**	**24.00**	**21.00**	**24.00**	**40.00**	**27.00**	**31.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	87.77	84.43	84.94	85.59	86.10	86.36	83.40	84.94	83.01	84.94
NORMAL. ANY HYP CORRECT	36.48	41.43	34.50	44.16	41.44	36.81	46.54	56.02	51.55	46.38
LANGUAGE IMPROVEMENT	26.38	26.13	23.94	26.26	26.38	26.26	24.84	23.94	23.04	24.58
SPEAKER IMPROVEMENT	31.02	27.67	27.54	28.96	29.73	29.47	26.77	27.29	27.29	28.32
ACTIVITY IMPROVEMENT	-2.83	-3.73	-1.41	-2.96	-2.57	-2.83	-3.09	-3.86	-3.34	-3.60
PLATFORM IMPROVEMENT	21.11	21.75	20.20	21.75	21.23	22.39	18.79	18.27	22.01	18.53
LSP IMPROVEMENT	57.78	54.31	54.82	55.59	56.11	56.11	53.15	54.56	52.89	54.82
ROUTING										
COMPLETELY CORRECT	97.55	97.43	97.81	97.43	97.68	98.20	97.43	97.55	95.75	97.68
PARTIALLY CORRECT	97.19	97.04	97.48	97.04	97.33	97.93	97.04	97.19	95.11	97.33
ACCURACY	97.19	97.04	97.48	97.04	97.33	97.93	97.04	97.19	95.11	97.33
EFFICIENCY	84.43	86.64	81.54	87.33	85.77	84.10	87.22	88.41	87.70	84.23
TOTAL	**634.07**	**630.14**	**618.84**	**638.19**	**636.53**	**632.73**	**629.13**	**641.50**	**630.12**	**630.54**
RANK	**27.00**	**31.00**	**41.00**	**23.00**	**25.00**	**29.00**	**33.00**	**18.00**	**32.00**	**30.00**
GRAND TOTAL	**1262.95**	**1260.84**	**1256.68**	**1256.22**	**1249.17**	**1252.58**	**1241.77**	**1222.57**	**1238.21**	**1230.58**
RANK FOR GRAND TOTAL	**18.00**	**19.00**	**20.00**	**21.00**	**25.00**	**24.00**	**26.00**	**30.00**	**27.00**	**28.00**
TOTAL RANK	**41.00**	**44.00**	**46.00**	**46.00**	**49.00**	**50.00**	**57.00**	**58.00**	**59.00**	**61.00**

SCENARIO GROUP (7,8,9)	100/50/75	25/100/0	100/25/50	50/100/50	75/100/75	50/25/75	75/50/75	75/100/100	100/75/100	75/25/75
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	25.78	22.78	27.13	25.65	26.76	27.83	26.57	25.34	27.33	27.39
SPEAKER IMPROVEMENT	33.21	26.89	30.97	22.81	25.38	24.09	25.38	26.25	28.42	26.97
ACTIVITY IMPROVEMENT	-0.24	0.00	-0.24	-0.49	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24
PLATFORM IMPROVEMENT	24.61	23.19	25.81	25.58	23.09	24.24	24.94	23.85	24.78	22.14
TOTAL	**83.36**	**72.86**	**83.67**	**73.54**	**74.99**	**75.92**	**76.65**	**75.20**	**80.29**	**76.27**
RANK	**23.00**	**45.00**	**18.00**	**43.00**	**42.00**	**38.00**	**34.00**	**41.00**	**31.00**	**36.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	55.20	58.06	48.21	55.55	55.58	57.56	52.64	53.98	46.52	53.48
NORMAL. ANY HYP CORRECT	29.88	31.87	25.34	32.87	33.23	32.82	35.35	36.90	35.75	30.47
LANGUAGE IMPROVEMENT	17.46	19.97	16.41	21.57	21.41	21.97	17.06	18.97	14.64	22.06
SPEAKER IMPROVEMENT	12.10	23.92	4.40	14.99	14.02	7.50	6.90	12.16	0.06	4.98
ACTIVITY IMPROVEMENT	-8.54	-4.39	-12.53	-5.37	-5.85	-7.70	-8.70	-7.62	-16.31	-14.64
PLATFORM IMPROVEMENT	17.03	21.70	15.60	22.13	17.79	19.23	16.52	19.05	12.97	16.98
LSP IMPROVEMENT	36.93	45.59	32.11	41.59	40.91	34.84	33.52	37.06	26.93	29.91
ROUTING										
COMPLETELY CORRECT	96.65	96.57	96.80	95.73	95.51	96.63	96.71	96.02	96.64	96.67
PARTIALLY CORRECT	96.12	96.01	96.29	95.04	94.79	96.08	96.18	95.37	96.10	96.13
ACCURACY	96.12	96.01	96.29	95.04	94.79	96.08	96.18	95.37	96.10	96.13
EFFICIENCY	60.42	60.79	51.48	62.58	63.86	60.99	64.05	63.87	64.73	63.03
TOTAL	**509.35**	**546.11**	**470.39**	**531.74**	**526.03**	**516.01**	**506.42**	**521.12**	**474.12**	**495.19**
RANK	**35.00**	**20.00**	**49.00**	**26.00**	**28.00**	**33.00**	**37.00**	**31.00**	**46.00**	**42.00**
GRAND TOTAL	**1092.86**	**1056.09**	**1056.09**	**1046.52**	**1050.99**	**1047.47**	**1042.94**	**1047.55**	**1036.12**	**1029.05**
RANK FOR GRAND TOTAL	**27.00**	**31.00**	**32.00**	**36.00**	**33.00**	**35.00**	**37.00**	**34.00**	**38.00**	**39.00**
TOTAL RANK	**58.00**	**65.00**	**67.00**	**69.00**	**70.00**	**71.00**	**71.00**	**72.00**	**77.00**	**78.00**

REPRODUCIBLE (1,2,3)	75/100/100	50/100/50	75/50/75	50/75/75	100/50/100	100/75/100	100/100/100	100/50/75	25/100/0	75/100/75
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	28.09	26.51	28.19	27.93	28.19	28.19	28.15	28.19	26.51	28.19
SPEAKER IMPROVEMENT	32.18	31.53	33.46	31.66	34.49	33.59	32.69	35.52	32.05	32.82
ACTIVITY IMPROVEMENT	0.00	0.00	0.00	0.00	-0.13	0.00	0.00	0.00	0.00	0.00
PLATFORM IMPROVEMENT	22.14	24.19	22.39	22.26	22.01	22.52	22.78	22.91	26.38	23.42
TOTAL	**82.41**	**82.23**	**84.04**	**81.85**	**84.56**	**84.30**	**83.62**	**86.62**	**84.94**	**84.43**
RANK	**41.00**	**43.00**	**37.00**	**46.00**	**33.00**	**36.00**	**39.00**	**29.00**	**32.00**	**35.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	83.78	84.30	83.91	83.40	82.37	82.24	82.24	80.95	79.28	80.31
NORMAL. ANY HYP CORRECT	57.36	56.73	53.97	57.20	45.32	49.54	58.23	37.12	47.43	49.93
LANGUAGE IMPROVEMENT	24.20	23.04	23.81	23.68	25.87	25.10	21.75	25.74	22.78	23.42
SPEAKER IMPROVEMENT	26.77	28.70	27.29	26.00	25.87	25.61	24.97	25.10	22.27	22.91
ACTIVITY IMPROVEMENT	-3.60	-3.47	-4.63	-3.99	-6.30	-6.18	-6.43	-6.95	-4.12	-4.24
PLATFORM IMPROVEMENT	18.27	19.56	17.63	18.27	18.79	18.79	16.09	18.92	21.62	18.02
LSP IMPROVEMENT	53.41	54.18	53.79	53.02	51.99	52.12	51.99	50.96	48.13	50.06
ROUTING										
COMPLETELY CORRECT	97.55	97.43	97.68	97.55	97.55	97.55	97.55	97.68	96.01	97.43
PARTIALLY CORRECT	97.19	97.04	97.33	97.19	97.19	97.19	97.19	97.33	95.41	97.04
ACCURACY	97.19	97.04	97.33	97.19	97.19	97.19	97.19	97.33	95.41	97.04
EFFICIENCY	88.29	86.18	86.33	89.13	83.99	84.75	85.75	81.92	87.14	85.40
TOTAL	**640.41**	**640.73**	**634.44**	**638.64**	**619.83**	**623.90**	**626.52**	**606.10**	**611.36**	**617.32**
RANK	**21.00**	**19.00**	**26.00**	**22.00**	**39.00**	**37.00**	**36.00**	**48.00**	**45.00**	**43.00**
GRAND TOTAL	**1217.28**	**1216.34**	**1222.72**	**1211.59**	**1211.75**	**1214.00**	**1211.86**	**1212.44**	**1205.94**	**1208.33**
RANK FOR GRAND TOTAL	**31.00**	**32.00**	**29.00**	**37.00**	**36.00**	**33.00**	**35.00**	**34.00**	**39.00**	**38.00**
TOTAL RANK	**62.00**	**62.00**	**63.00**	**68.00**	**72.00**	**73.00**	**75.00**	**77.00**	**77.00**	**78.00**

SCENARIO GROUP (7,8,9)	50/75/50	25/75/25	100/25/75	50/75/100	75/75/100	50/75/75	50/50/75	100/50/100	100/100/100	75/50/100
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	25.58	25.54	28.27	23.68	26.96	26.97	25.09	26.20	26.09	27.39
SPEAKER IMPROVEMENT	25.78	20.03	29.65	20.98	21.54	20.77	24.05	28.11	26.85	23.58
ACTIVITY IMPROVEMENT	-0.24	-1.47	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24
PLATFORM IMPROVEMENT	24.53	25.77	21.78	22.06	21.19	22.90	23.86	24.61	23.65	24.63
TOTAL	**75.65**	**69.87**	**79.45**	**66.48**	**69.45**	**70.41**	**72.76**	**78.68**	**76.35**	**75.36**
RANK	**39.00**	**51.00**	**32.00**	**56.00**	**52.00**	**49.00**	**46.00**	**33.00**	**35.00**	**40.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	49.02	52.31	46.95	57.07	54.02	50.51	47.47	40.27	40.93	44.02
NORMAL. ANY HYP CORRECT	32.29	30.61	30.15	37.94	39.07	35.99	35.98	32.26	33.20	32.98
LANGUAGE IMPROVEMENT	18.47	22.43	16.12	19.23	18.39	19.64	14.45	11.86	14.20	16.44
SPEAKER IMPROVEMENT	7.36	13.65	0.46	8.59	8.41	5.41	-2.09	-3.99	-5.06	-5.81
ACTIVITY IMPROVEMENT	-9.69	-5.98	-15.86	-4.24	-7.48	-6.96	-10.83	-21.11	-19.19	-17.72
PLATFORM IMPROVEMENT	18.75	22.10	9.52	17.87	13.82	16.98	13.65	12.14	12.88	13.89
LSP IMPROVEMENT	32.87	40.32	27.16	34.23	33.89	30.31	24.03	20.92	20.32	21.65
ROUTING										
COMPLETELY CORRECT	96.33	95.33	96.73	96.13	95.55	96.05	96.49	96.58	95.56	96.73
PARTIALLY CORRECT	95.74	94.57	96.20	95.50	94.83	95.41	95.93	96.03	94.84	96.20
ACCURACY	95.74	94.57	96.20	95.50	94.83	95.41	95.93	96.03	94.84	96.20
EFFICIENCY	62.48	61.39	58.20	66.37	64.37	63.76	63.59	61.37	62.74	61.88
TOTAL	**499.36**	**521.30**	**461.82**	**524.18**	**509.69**	**502.50**	**474.62**	**442.36**	**445.27**	**456.44**
RANK	**39.00**	**30.00**	**53.00**	**29.00**	**34.00**	**38.00**	**44.00**	**61.00**	**60.00**	**57.00**
GRAND TOTAL	**1028.94**	**1010.36**	**1017.98**	**989.52**	**995.85**	**995.35**	**983.92**	**993.11**	**979.73**	**983.96**
RANK FOR GRAND TOTAL	**40.00**	**42.00**	**41.00**	**46.00**	**43.00**	**44.00**	**48.00**	**45.00**	**49.00**	**47.00**
TOTAL RANK	**78.00**	**81.00**	**85.00**	**85.00**	**86.00**	**87.00**	**90.00**	**94.00**	**95.00**	**97.00**

REPRODUCIBLE (1,2,3)	50/50/75	100/25/50	50/100/100	100/25/75	25/75/25	75/25/50	75/75/100	50/75/100	75/50/100	50/100/75
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	28.19	28.06	26.51	28.06	26.51	28.06	27.93	28.19	28.19	26.51
SPEAKER IMPROVEMENT	32.05	36.30	30.76	34.88	29.09	35.27	30.89	30.37	32.31	29.47
ACTIVITY IMPROVEMENT	0.00	-0.13	0.00	-0.13	0.00	-0.13	0.00	0.00	-0.13	0.00
PLATFORM IMPROVEMENT	21.88	24.19	22.14	21.75	25.22	22.65	22.14	21.62	21.88	22.91
TOTAL	**82.12**	**88.42**	**79.41**	**84.56**	**80.82**	**85.85**	**80.96**	**80.18**	**82.25**	**78.89**
RANK	**44.00**	**22.00**	**56.00**	**33.00**	**50.00**	**30.00**	**49.00**	**54.00**	**42.00**	**57.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	81.85	80.95	82.63	81.60	83.14	78.12	79.92	81.21	78.12	79.54
NORMAL. ANY HYP CORRECT	49.61	26.55	59.84	37.70	50.72	30.11	50.57	50.49	42.05	55.06
LANGUAGE IMPROVEMENT	25.10	26.51	22.65	26.38	24.20	27.03	24.97	24.33	25.48	21.75
SPEAKER IMPROVEMENT	24.71	24.20	26.77	26.39	27.16	23.17	22.91	23.94	21.49	22.52
ACTIVITY IMPROVEMENT	-3.86	-7.08	-3.73	-7.08	-2.31	-8.75	-6.43	-4.76	-9.01	-4.50
PLATFORM IMPROVEMENT	19.18	19.95	17.89	19.05	20.98	18.92	18.40	18.53	18.66	17.50
LSP IMPROVEMENT	51.35	50.83	52.12	51.35	52.89	48.00	49.67	50.70	47.61	49.42
ROUTING										
COMPLETELY CORRECT	97.68	98.20	97.55	97.68	95.62	98.07	97.55	97.43	98.20	97.43
PARTIALLY CORRECT	97.33	97.93	97.19	97.33	94.96	97.78	97.19	97.04	97.93	97.04
ACCURACY	97.33	97.93	97.19	97.33	94.96	97.78	97.19	97.04	97.93	97.04
EFFICIENCY	87.25	62.89	86.89	78.21	85.81	69.33	86.32	87.57	84.10	86.99
TOTAL	**627.53**	**578.86**	**636.99**	**605.94**	**628.13**	**579.56**	**618.26**	**623.52**	**602.56**	**619.79**
RANK	**35.00**	**58.00**	**24.00**	**49.00**	**34.00**	**57.00**	**42.00**	**38.00**	**51.00**	**40.00**
GRAND TOTAL	**1202.37**	**1197.80**	**1192.86**	**1197.86**	**1193.87**	**1180.51**	**1184.98**	**1184.78**	**1178.31**	**1172.02**
RANK FOR GRAND TOTAL	**40.00**	**42.00**	**44.00**	**41.00**	**43.00**	**47.00**	**45.00**	**46.00**	**48.00**	**49.00**
TOTAL RANK	**79.00**	**80.00**	**80.00**	**82.00**	**84.00**	**87.00**	**91.00**	**92.00**	**93.00**	**97.00**

SCENARIO GROUP (7,8,9)	25/100/25	75/25/100	100/25/100	50/100/75	50/100/100	25/75/75	50/50/100	25/50/100	25/75/50	25/50/75
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	23.66	26.78	25.65	25.89	26.49	24.71	25.81	25.49	24.44	25.83
SPEAKER IMPROVEMENT	16.98	22.15	22.66	19.00	21.62	18.81	19.44	15.40	15.56	14.99
ACTIVITY IMPROVEMENT	-1.47	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24	-0.24
PLATFORM IMPROVEMENT	24.15	22.07	24.96	24.17	23.11	21.54	24.00	21.69	23.02	22.39
TOTAL	**63.33**	**70.75**	**73.03**	**68.82**	**70.99**	**64.83**	**69.00**	**62.35**	**62.78**	**62.97**
RANK	**61.00**	**48.00**	**44.00**	**54.00**	**47.00**	**58.00**	**53.00**	**64.00**	**63.00**	**62.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	49.31	48.52	47.87	43.31	41.41	42.31	42.32	45.37	46.62	43.47
NORMAL. ANY HYP CORRECT	27.72	30.59	28.16	36.68	35.05	29.92	34.30	33.13	32.99	32.49
LANGUAGE IMPROVEMENT	21.69	17.11	17.10	17.49	11.31	20.84	15.52	16.67	16.34	15.63
SPEAKER IMPROVEMENT	10.34	-1.81	-1.50	1.41	-4.63	-1.16	-4.63	0.39	-0.61	-1.46
ACTIVITY IMPROVEMENT	-5.54	-13.15	-19.12	-15.93	-15.65	-6.92	-16.34	-9.89	-8.73	-11.22
PLATFORM IMPROVEMENT	22.05	12.12	16.70	15.53	9.35	17.98	12.95	14.13	16.60	14.20
LSP IMPROVEMENT	35.57	25.04	23.19	28.00	20.36	25.39	21.87	23.69	24.30	24.18
ROUTING										
COMPLETELY CORRECT	95.62	96.65	96.69	94.91	95.88	94.13	96.58	96.26	94.02	94.93
PARTIALLY CORRECT	94.92	96.11	96.16	94.09	95.21	93.18	96.03	95.66	93.05	94.10
ACCURACY	94.92	96.11	96.16	94.09	95.21	93.18	96.03	95.66	93.05	94.10
EFFICIENCY	60.83	57.60	55.65	64.08	63.68	62.56	65.01	63.07	61.74	63.94
TOTAL	**507.43**	**464.89**	**457.05**	**473.66**	**447.19**	**471.40**	**459.62**	**474.13**	**469.36**	**464.36**
RANK	**36.00**	**51.00**	**56.00**	**47.00**	**59.00**	**48.00**	**55.00**	**45.00**	**50.00**	**52.00**
GRAND TOTAL	**950.74**	**960.17**	**968.26**	**955.38**	**944.10**	**925.18**	**942.62**	**910.55**	**908.80**	**905.13**
RANK FOR GRAND TOTAL	**53.00**	**51.00**	**50.00**	**52.00**	**54.00**	**56.00**	**55.00**	**57.00**	**58.00**	**59.00**
TOTAL RANK	97.00	99.00	100.00	101.00	106.00	106.00	108.00	109.00	113.00	114.00

REPRODUCIBLE (1,2,3)	75/25/100	25/100/25	100/25/100	25/100/100	25/75/50	75/25/75	25/75/75	50/25/75	50/50/100	25/50/100
MERGE ALGORITHM										
LANGUAGE IMPROVEMENT	28.06	26.51	27.93	26.51	26.51	28.19	26.51	28.19	27.80	26.51
SPEAKER IMPROVEMENT	32.69	28.32	34.62	27.41	26.77	32.31	28.19	32.18	31.28	25.36
ACTIVITY IMPROVEMENT	-0.13	0.00	-0.13	0.00	0.00	-0.13	0.00	-0.13	-0.13	-0.13
PLATFORM IMPROVEMENT	21.11	25.61	21.36	21.75	22.78	21.62	21.88	21.36	21.36	21.36
TOTAL	**81.73**	**80.44**	**83.78**	**75.67**	**76.06**	**81.99**	**76.58**	**81.60**	**80.31**	**73.10**
RANK	**47.00**	**51.00**	**38.00**	**62.00**	**60.00**	**45.00**	**59.00**	**48.00**	**52.00**	**66.00**
CORRELATION ALGORITHM										
TOP HYPOTHESIS CORRECT	81.60	74.26	50.71	79.67	78.89	73.36	75.80	75.80	77.48	78.38
NORMAL. ANY HYP CORRECT	31.03	47.47	22.85	52.39	55.58	28.13	51.99	29.30	27.39	48.32
LANGUAGE IMPROVEMENT	27.80	22.39	24.58	22.52	20.21	25.74	21.11	26.38	26.90	23.30
SPEAKER IMPROVEMENT	25.48	17.38	-0.90	22.78	22.27	19.31	18.41	20.98	20.98	21.24
ACTIVITY IMPROVEMENT	-7.85	-4.37	-35.52	-3.99	-6.18	-14.03	-4.76	-12.10	-10.94	-6.18
PLATFORM IMPROVEMENT	19.18	20.08	14.41	17.50	15.44	18.02	15.83	17.63	19.05	17.24
LSP IMPROVEMENT	51.35	43.50	22.39	48.64	48.51	43.24	44.91	45.55	47.10	47.87
ROUTING										
COMPLETELY CORRECT	97.68	95.75	98.58	95.62	95.75	97.94	97.30	97.94	98.07	97.30
PARTIALLY CORRECT	97.33	95.11	98.37	94.96	95.11	97.63	96.89	97.63	97.78	96.89
ACCURACY	97.33	95.11	98.37	94.96	95.11	97.63	96.89	97.63	97.78	96.89
EFFICIENCY	70.49	85.83	44.12	86.74	86.64	67.66	86.17	71.09	70.36	86.97
TOTAL	**591.42**	**592.52**	**437.96**	**611.79**	**607.33**	**554.63**	**600.54**	**567.83**	**571.95**	**608.22**
RANK	**54.00**	**53.00**	**66.00**	**44.00**	**47.00**	**65.00**	**52.00**	**64.00**	**60.00**	**46.00**
GRAND TOTAL	**1163.53**	**1155.62**	**1024.42**	**1141.48**	**1139.75**	**1128.56**	**1136.60**	**1139.03**	**1134.12**	**1119.92**
RANK FOR GRAND TOTAL	**50.00**	**51.00**	**66.00**	**52.00**	**53.00**	**59.00**	**55.00**	**54.00**	**56.00**	**62.00**
TOTAL RANK	101.00	104.00	104.00	106.00	107.00	110.00	111.00	112.00	112.00	112.00

SCENARIO GROUP (7,8,9)	25/25/75	25/100/75	50/25/100	25/100/100	25/100/50	25/25/100	25/75/100
MERGE ALGORITHM							
LANGUAGE IMPROVEMENT	25.03	22.61	27.60	23.92	25.01	24.96	24.39
SPEAKER IMPROVEMENT	20.92	15.80	18.45	18.29	15.64	15.83	15.15
ACTIVITY IMPROVEMENT	-0.24	-0.24	-0.24	-0.24	-0.49	-0.24	-0.24
PLATFORM IMPROVEMENT	24.69	22.61	22.19	21.62	24.00	25.05	21.35
TOTAL	**70.40**	**60.78**	**68.00**	**63.59**	**64.15**	**65.61**	**60.64**
RANK	**50.00**	**65.00**	**55.00**	**60.00**	**59.00**	**57.00**	**66.00**
CORRELATION ALGORITHM							
TOP HYPOTHESIS CORRECT	33.42	42.25	31.32	36.99	36.72	28.80	39.76
NORMAL. ANY HYP CORRECT	28.29	29.63	26.83	29.25	30.13	37.52	29.86
LANGUAGE IMPROVEMENT	10.81	15.59	11.70	15.31	13.20	-13.15	18.36
SPEAKER IMPROVEMENT	-16.15	-0.45	-19.12	-7.28	-6.76	-24.42	-3.41
ACTIVITY IMPROVEMENT	-24.93	-5.54	-26.14	-9.51	-14.77	-36.59	-8.82
PLATFORM IMPROVEMENT	12.50	16.68	6.63	14.56	11.90	-14.84	14.94
LSP IMPROVEMENT	9.17	22.65	7.45	15.27	20.23	3.34	20.74
ROUTING							
COMPLETELY CORRECT	96.45	93.74	96.57	94.79	93.43	93.91	94.49
PARTIALLY CORRECT	95.87	92.72	96.01	93.95	92.37	92.92	93.59
ACCURACY	95.87	92.72	96.01	93.95	92.37	92.92	93.59
EFFICIENCY	59.22	60.61	55.28	60.36	59.45	62.26	62.74
TOTAL	**400.51**	**460.59**	**382.56**	**437.64**	**428.27**	**322.67**	**455.84**
RANK	**64.00**	**54.00**	**65.00**	**62.00**	**63.00**	**66.00**	**58.00**
GRAND TOTAL	**893.34**	**886.02**	**858.53**	**882.75**	**877.35**	**781.91**	**880.32**
RANK FOR GRAND TOTAL	**60.00**	**61.00**	**65.00**	**62.00**	**64.00**	**66.00**	**63.00**
TOTAL RANK	**114.00**	**119.00**	**120.00**	**122.00**	**122.00**	**123.00**	**124.00**

REPRODUCIBLE (1,2,3)	50/25/100	25/100/50	25/50/75	25/25/75	25/25/100	25/100/75	25/75/100
MERGE ALGORITHM							
LANGUAGE IMPROVEMENT	27.80	26.51	26.51	28.19	27.67	26.51	26.51
SPEAKER IMPROVEMENT	31.28	26.77	27.16	30.50	26.00	27.41	27.29
ACTIVITY IMPROVEMENT	-0.13	0.00	0.00	-0.13	-0.13	0.00	0.00
PLATFORM IMPROVEMENT	21.36	23.55	21.49	21.11	20.59	22.14	21.49
TOTAL	**80.31**	**76.83**	**75.16**	**79.67**	**74.13**	**76.06**	**75.29**
RANK	**52.00**	**58.00**	**64.00**	**55.00**	**65.00**	**60.00**	**63.00**
CORRELATION ALGORITHM							
TOP HYPOTHESIS CORRECT	77.48	72.72	77.48	75.42	74.77	71.94	72.84
NORMAL. ANY HYP CORRECT	27.39	47.84	46.97	30.62	63.63	48.37	45.61
LANGUAGE IMPROVEMENT	26.90	22.65	22.52	26.51	13.90	19.05	21.75
SPEAKER IMPROVEMENT	20.98	16.35	21.11	17.50	16.60	14.42	16.22
ACTIVITY IMPROVEMENT	-10.94	-3.73	-6.69	-13.12	-14.41	-8.23	-9.26
PLATFORM IMPROVEMENT	19.05	18.40	17.50	19.18	6.69	13.26	15.96
LSP IMPROVEMENT	47.10	42.21	46.84	44.91	44.14	41.05	42.08
ROUTING							
COMPLETELY CORRECT	98.07	95.62	97.43	98.20	97.30	95.62	95.62
PARTIALLY CORRECT	97.78	94.96	97.04	97.93	96.89	94.96	94.96
ACCURACY	97.78	94.96	97.04	97.93	96.89	94.96	94.96
EFFICIENCY	70.36	86.97	85.62	74.52	86.05	86.04	84.68
TOTAL	**571.95**	**588.95**	**602.86**	**569.60**	**582.48**	**571.44**	**575.42**
RANK	**61.00**	**55.00**	**50.00**	**63.00**	**56.00**	**62.00**	**59.00**
GRAND TOTAL	**1134.12**	**1126.76**	**1128.98**	**1127.29**	**1101.36**	**1103.86**	**1102.45**
RANK FOR GRAND TOTAL	**56.00**	**61.00**	**58.00**	**60.00**	**65.00**	**63.00**	**64.00**
TOTAL RANK	**113.00**	**113.00**	**114.00**	**118.00**	**121.00**	**122.00**	**122.00**

APPENDIX H

TABLE WITH ALL RESULTS RANKED

SCENARIO GROUP 1,2,3	75/100/25	100/100/25	75/50/0	75/100/0	75/75/25	75/100/50	50/75/0
MERGE ALGORITHM							
LANGUAGE IMPROVEMENT	27.80	27.76	28.65	28.61	28.96	27.37	27.69
SPEAKER IMPROVEMENT	35.59	36.35	36.77	36.48	36.20	35.34	35.60
ACTIVITY IMPROVEMENT	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PLATFORM IMPROVEMENT	26.88	26.60	26.81	26.27	26.06	27.72	26.09
TOTAL	**90.26**	**90.71**	**92.23**	**91.36**	**91.22**	**90.42**	**89.37**
RANK	**15.00**	**12.00**	**4.00**	**8.00**	**9.00**	**14.00**	**20.00**
CORRELATION ALGORITHM							
TOP HYPOTHESIS CORRECT	90.39	88.08	88.48	86.26	86.58	86.73	86.96
NORMALIZ. ANY HYP CORRECT	55.59	50.92	41.29	49.39	49.46	50.35	52.20
TOTAL	**145.98**	**139.01**	**129.77**	**135.65**	**136.04**	**137.08**	**139.16**
RANK	**1.00**	**9.00**	**30.00**	**15.00**	**14.00**	**12.00**	**8.00**
LANGUAGE IMPROVEMENT	25.97	25.04	25.77	24.96	25.69	25.19	23.96
SPEAKER IMPROVEMENT	33.46	32.09	31.51	29.88	29.85	31.08	30.84
ACTIVITY IMPROVEMENT	-1.68	-2.01	-2.16	-2.83	-2.43	-1.84	-2.76
PLATFORM IMPROVEMENT	24.24	23.27	23.41	22.47	22.12	24.38	22.05
LSP IMPROVEMENT	61.67	59.74	59.98	57.52	57.95	58.90	57.74
TOTAL	**143.65**	**138.14**	**138.51**	**131.99**	**133.19**	**137.71**	**131.82**
RANK	**1.00**	**4.00**	**3.00**	**14.00**	**13.00**	**5.00**	**15.00**
ROUTING							
COMPLETELY CORRECT	97.43	97.44	97.55	97.55	97.55	97.55	97.56
PARTIALLY CORRECT	97.04	97.05	97.19	97.19	97.19	97.19	97.20
ACCURACY	97.04	97.05	97.19	97.19	97.19	97.19	97.20
EFFICIENCY	89.94	89.31	89.80	89.59	89.32	88.41	89.56
TOTAL	**381.45**	**380.84**	**381.73**	**381.52**	**381.25**	**380.34**	**381.53**
RANK	**5.00**	**11.00**	**1.00**	**3.00**	**7.00**	**14.00**	**2.00**
GRAND TOTAL	**761.33**	**748.69**	**742.24**	**740.52**	**741.69**	**745.56**	**741.89**
RANK FOR GRAND TOTAL	**1.00**	**2.00**	**6.00**	**9.00**	**8.00**	**4.00**	**7.00**
TOTAL RANK	**22.00**	**36.00**	**38.00**	**40.00**	**43.00**	**45.00**	**45.00**

☆U.S. GOVERNMENT PRINTING OFFICE: 1995-610-126-20031

Rome Laboratory

Customer Satisfaction Survey

RL-TR-________

Please complete this survey, and mail to RL/IMPS, 26 Electronic Pky, Griffiss AFB NY 13441-4514. Your assessment and feedback regarding this technical report will allow Rome Laboratory to have a vehicle to continuously improve our methods of research, publication, and customer satisfaction. Your assistance is greatly appreciated.
Thank You

Organization Name:______________________________________(Optional)

Organization POC: ______________________________________(Optional)

Address:____

1. On a scale of 1 to 5 how would you rate the technology developed under this research?

5-Extremely Useful 1-Not Useful/Wasteful

Rating_____

Please use the space below to comment on your rating. Please suggest improvements. Use the back of this sheet if necessary.

2. Do any specific areas of the report stand out as exceptional?

Yes___ No______

If yes, please identify the area(s), and comment on what aspects make them "stand out."

3. Do any specific areas of the report stand out as inferior?

Yes___ No___

If yes, please identify the area(s), and comment on what aspects make them "stand out."

4. Please utilize the space below to comment on any other aspects of the report. Comments on both technical content and reporting format are desired.